VOLUME 26 MAY, 1938 NUMBER 5

PROCEEDINGS
of
The Institute of Radio Engineers

Thirteenth Annual Convention
June 16, 17, and 18, 1938
New York, N. Y.

Application Blank for Associate Membership on Page XIII

Institute of Radio Engineers Forthcoming Meetings

THIRTEENTH
ANNUAL CONVENTION
New York, N. Y.
June 16, 17, and 18, 1938

BUFFALO-NIAGARA SECTION
May 18, 1938

CHICAGO SECTION
May 6, 1938

CINCINNATI SECTION
May 24, 1938

MONTREAL SECTION
May 11, 1938

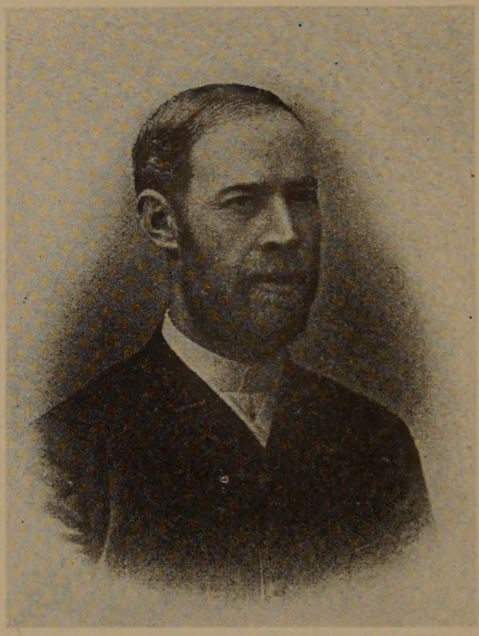
NEW ORLEANS SECTION
May 13, 1938

PHILADELPHIA SECTION
May 6, 1938

PITTSBURGH SECTION
May 17, 1938
June 21, 1938

TORONTO SECTION
May 9, 1938

WASHINGTON SECTION
May 9, 1938



Courtesy American Institute of Physics

Heinrich Hertz

1857-1894

INSTITUTE NEWS AND RADIO NOTES

April Meeting of the Board of Directors

The April meeting of the Board of Directors was held on the 6th at the Institute office. Those present were C. M. Jansky, Jr., acting chairman; Melville Eastham, treasurer; H. H. Beverage, Ralph Bown, J. E. Brown, F. W. Cunningham, Virgil M. Graham, R. A. Hackbusch, L. C. F. Horle, B. J. Thompson, H. M. Turner, and H. P. Westman, secretary.

R. M. Wilmotte was transferred to Fellow grade, and Noel Ashbridge, Harold Bishop, W. F. Einthoven, and L. W. Hayes were admitted to that grade. The following were transferred to Member grade: P. C. Goldmark, E. W. Herold, H. E. Hiller, W. M. Sharpless, and W. G. Webster. Admission to Member grade was granted to W. J. Andrews, R. E. Coram, V. H. Dudman, R. D. Lawlor, W. L. McPherson, H. A. Reise, R. H. Scott, F. C. Stockwell, and F. D. Webster.

Sixty-six Associates, one Junior, and twenty-one Students were elected to membership.

An invitation to the Institute to be listed as an organization approving and endorsing the Model Law for the Registration of Professional Engineers and Land Surveyors was declined on the basis of a statement of policy previously adopted, which reads as follows:

"As an enunciation of Board policy, it was stated that the Board sees no need in the public interest for the governmental licensing of practicing radio engineers, and further urges that if such licensing is nevertheless enforced, the requirements for such license shall be reasonably based on the qualifications for practicing radio engineering exclusively."

Committee Work

Admissions Committee

A meeting of the Admissions Committee, which was attended by F. W. Cunningham, chairman; Melville Eastham, J. F. Farrington, L. C. F. Horle, C. M. Jansky, Jr., E. R. Shute, A. F. Van Dyck, and H. P. Westman, secretary, was held on April 6 in the Institute office. An application for transfer to the grade of Fellow was reconsidered and the previous decision reaffirmed. Four out of five applications for transfer to Member grade were approved and three applications for transfer to Member grade were accepted. A special application for Associate membership was considered and approved.

Awards Committee

The Awards Committee met in the Institute office on April 12. Those present were H. H. Beverage, chairman; Alan Hazeltine, H. M. Turner, and H. P. Westman, secretary. Recommendations were prepared for recipients for the Institute Medal of Honor, the Morris Liebmann Memorial Prize, and the Proceedings Paper Prize.

Membership Committee

The Membership Committee met in the Institute office on April 6. Those present were C. E. Scholz, chairman; H. A. Chinn, J. M. Clayton, I. S. Coggeshall, F. W. Cunningham, E. T. Dickey (representing F. X. Rettenmeyer), Coke Flannagan, R. M. Heintz, L. G. Pacent, Bernard Salzberg, and J. D. Crawford, assistant secretary. The committee spent most of its time in discussing the existing distribution of membership among the various grades and the qualifications required for these grades of membership.

New York Program Committee

The New York Program Committee met in the Institute office on March 24 and those present were Austin Bailey, chairman; A. B. Chamberlain, Keith Henney, G. T. Royden, and H. P. Westman, secretary. Arrangements were made for the May New York meeting.

New York Registration of Engineers Committee

On April 4, the committee, established to investigate the situation brought about through the present New York state law for the registration of engineers, met in the Institute office. Those present were L. M. Hull, chairman; L. C. F. Horle, E. L. Nelson, A. F. Van Dyck, and H. P. Westman, secretary.

Institute Meetings

Atlanta

The Atlanta Section met on March 17 at the Atlanta Athletic Club. There were twenty present, and C. F. Daugherty, chairman, presided.

"A Report of the Broadcast Engineering Conference held at the Ohio State University," was presented by Ben Akerman, chief engineer of WGST. He outlined first the program of the conference, then presented a brief résumé of subjects covered at it. Each of the lectures presented was summarized. A description was then given of an inspection trip to the transmitting plant of WLW.

Buffalo-Niagara

A meeting of the Buffalo-Niagara Section was held on March 9 at the University of Buffalo, and attended by thirty-five. G. C. Crom, Jr., chairman, presided.

H. C. Tittle, chief engineer of the Colonial Radio Corporation, presented a paper on "Spurious Responses in Radio Receivers." The effects of various spurious responses were described and their sources and methods used in their elimination or suppression covered.

The April meeting of the section was held on the 4th at the Hotel Statler and was attended by ninety. Chairman Crom presided.

H. H. Scott of the General Radio Company presented a paper on "Stroboscopes and High-Speed Photography." Applications of the stroboscope including its use with high-speed cameras were described. They involved automotive, Diesel-engine, aviation, and other industries and included effects of strains, crankshaft whip, injection jets, speeds of textile spindles, and testing of materials. Their application to the photoelectric integraph at the Massachusetts Institute of Technology, was described. High-speed motion pictures illustrated in slow motion details of many phenomena of everyday experience which cannot be followed by the eye.

The paper was closed with a discussion of the application of a photoelectric cell in a color comparator. The use of sound-level-measuring devices was also covered.

Chicago

The April 1 meeting of the Chicago Section was held jointly with the Western Society of Engineers and the local sections of the American Institute of Electrical Engineers and the Society of Motion Picture Engineers. There were five hundred present and J. E. Brown, chairman, presided.

R. R. Beal, research director of the Radio Corporation of America, presented a paper on "The RCA High-Definition Television System." He described the television facilities provided for the present field tests in the New York metropolitan area. There were also reviewed the elements of the iconoscope, methods of analyzing, transmitting, and resynthesizing the scene to be transmitted, and the transmitting and receiving equipment.

Cincinnati

G. F. Platts presided in the absence of Chairman Rockwell at the March 15 meeting of the Cincinnati Section. There were fifty-two present and the meeting was held at the University of Cincinnati.

Philip Konkle, design engineer at stations WLW and WSAI, presented a paper on "A Review of Modern Television Principles." He described first the older systems using mechanical scanning. He then outlined the band widths necessary for successful transmission when using systems transmitting various numbers of lines. It was stated that improved resolution resulted from interlaced scanning. He then described such subjects as flicker, the generation of saw-tooth wave forms for magnetic deflection for cathode-ray tubes, the characteristics of the iconoscope, and illumination problems. The paper was closed with a discussion of economic factors, program material, and the commercial development of television.

Cleveland

On November 30, a meeting of the Cleveland Section was held at WHK studios and attended by fifty-seven. R. A. Fox, chairman, presided.

"Matrices in Electrical Engineering Problems" was the subject of a paper by R. S. Burington, professor of electrical engineering at Case School of Applied Science. Dr. Burington started with simple explanations of algebraic formulas and how these equations may be converted into matrices. It was shown that any algebraic problem can be converted into a matrix. Once the terms of the equation have been so arranged, any solution of the problem can be obtained quickly. Electrical meshes involving complex and interconnected elements can be converted to matrices and solutions for current, power, impedance, and other unknowns may easily be found. The use of matrices avoids long calculations and materially simplifies electrical engineering problems.

The second paper of the evening was on "Standing Waves in Broadcast Studios," by E. L. Gove, technical supervisor of WHK and WCLE. He presented first a brief discussion of the phenomena encountered and proceeded with the use of two microphones and a sound source to demonstrate cancellation and reinforcement of sound waves caused by the acoustical boundaries provided by the room.

The annual meeting of the section was held on January 27 in the Ohio Bell Telephone Company auditorium. There were thirty-one present and R. A. Fox, chairman, presided. The evening was devoted to a tour of the telephone plant which included long-distance operation, broadcast-network facilities, and automatic-exchange equipment. Sound-motion-picture film was shown and covered the early telephone equipment and its development to date.

In the election of officers, L. N. Chatterton of the department of

public safety transmitting station WRBH, was named chairman. G. E. Grostick of Bud Radio, Inc., was elected vice chairman, and R. L. Kline of Winteradio, Inc., was re-elected secretary-treasurer.

L. N. Chatterton, chairman, presided at the February 18 meeting which was held jointly with the local section of the American Institute of Electrical Engineers. There were three hundred and fifty present at the WTAM studios in which the meeting was held.

S. E. Leonard, chief engineer of WTAM introduced O. B. Hanson, vice president and chief engineer of the National Broadcasting Corporation and several other members of the National Broadcasting Corporation staff who described the design and construction of the WTAM studios. This was followed by an inspection tour of the studios. Both iconoscope and kinescope tubes were on display as was the latest portable field equipment.

The March 24 meeting was held in the Hotel Statler with Chairman Chatterton presiding and twenty-nine in attendance.

"Coast Guard Activities" was the subject of a paper by M. H. Dunbar, radioelectrician in the Cleveland area of the U. S. Coast Guard. There were presented first motion pictures covering various phases of coast guard activities, including training of personnel. It was pointed out that admission to coast guard cadet schools was on a competitive and not an appointive basis and only about one per cent of those who applied during 1937 were accepted. Communication trucks were then described. They carry 150-watt transmitters as well as portable equipment useful for emergency service. The transmitters can be operated while the trucks are in motion. The trucks are of three-ton capacity and capable of traveling at fifty-five miles an hour.

Connecticut Valley Section

On February 10 a meeting of the Connecticut Valley Section was held in the Building of Electrical Progress, Springfield, Mass. There were thirty present and D. E. Noble, chairman, presided.

P. Robinson, engineer for Sprague Specialties, presented a paper on "Electrolytic Condensers." In it, Dr. Robinson pointed out that the equivalent circuit of a wet electrolytic condenser was two capacitances in series, the anode capacitance and the electrolyte capacitance, which are shunted by two resistances in series. One of the resistances, that of the anode film, does not obey Ohm's law and may have any value from 20,000 ohms to 20 megohms. The other is the electrolyte resistance, of the order of 100 ohms, and is not in evidence at low frequencies. On the basis of this equivalent network the temperature and frequency

characteristics of the device were described. Electromechanical processes were covered and the gain obtained by etching the metals employed, discussed. The application of dry electrolytic condensers for by-pass purposes at high frequencies was described.

On March 18, Chairman Noble presided at a meeting of the Connecticut Valley Section held in the Hartford Electric Light Company auditorium and attended by thirty.

Stanford Goldman, engineer for the General Electric Company at Bridgeport, Connecticut, presented a paper on "Regeneration in the Cathode Circuit of Tuned Amplifiers." Dr. Goldman presented first equations for the regeneration voltages in terms of circuit parameters. It was pointed out that because of these regenerative currents, the drop across the cathode resistor may be as great as ten times that caused by the plate current alone. A graphical method for determining the effects of regeneration was demonstrated.

Detroit

On March 2, the Detroit Section attended a meeting sponsored by the Engineering Society of Detroit and the local section of the American Institute of Electrical Engineers at the Masonic Temple. There were about 3000 present.

A paper on "Television—Its Problems and Progress," was the subject of a paper by A. F. Murray, engineer in charge of television research for the Philco Radio and Television Corporation. By the use of lantern slides, tubes, and receiving apparatus, the subject was presented from both the engineering and general public point of view. The effect on television pictures of interference from automobile ignition systems was demonstrated and indicated that a car may interfere with reception when it is one hundred feet from the receiving antenna. When the ignition system is properly filtered, the car may come within a few feet of the antenna without causing trouble. Systems in existence in England, France, Germany, and Japan were discussed. The German system employs a motion-picture camera mounted on the roof of a truck. After exposure the film passes through the roof into the interior of the truck where it is developed and, while still wet, scanned for television transmission. This procedure requires about twenty-six seconds. At the close of the talk, the stage curtain was raised to reveal a scene of a television broadcast station of the year 1940. Charles Penman, WJR announcer, played the part of the announcer and introduced Alex Dow, president of the Detroit Edison Company, who closed the meeting with a short address.

On March 25, the Detroit Section met in the studios of WWJ. E. H. I. Lee, chairman, presided.

H. J. Lyman, of the engineering department of the Philco Radio and Television Corporation, presented a paper on "Measurements of Characteristics of Automobile Antennas." A portable 550-kilocycle oscillator is connected to the antenna to be measured and a variable condenser adjusted to obtain the original frequency. The antenna capacitance is the difference between the two values of the condenser. A resonant bridge is used for detecting small departures in the oscillator frequency from the original value. When the capacitance and effective height of antennas are measured the inductance can be disregarded in the range over which the receivers operate. A standard antenna calibrated in terms of effective height permits a quick comparison of new antennas and the measurement of their characteristics. It has been found that the capacitance of a roof antenna installed in a fabric top ranges between 90 micromicrofarads and 700 micromicrofarads. The insulated steel-top antenna has a very low effective height and a capacitance range of between 1500 to 3000 micromicrofarads. Its power factor is poor and may run as high as 8 per cent. The undercarriage antenna has a capacitance of 100 to 250 micromicrofarads. The monorod or over-the-top antennas have capacitances as low as 15 to 40 micromicrofarads and the lead-in capacitance is larger than that of the antenna proper.

Los Angeles

On February 24, the Los Angeles Section met jointly with the local sections of the American Institute of Electrical Engineers and the Society of Motion Picture Engineers in the Southern California Edison Building auditorium. There were thirty-five present.

A paper on "What Mathematics is Doing for Electrical Communication" was the subject of a paper presented by T. C. Fry of the Bell Telephone Laboratories.

On March 15, sixty-eight members of the section met at the Los Angeles Junior College auditorium with R. O. Brooke, chairman, presiding.

A symposium on "Aviation Radio" was presented. Delmar Wright, an engineer for Bendix Aviation Radio, presented a paper on "Instrument Landing Systems." He covered methods by which landing beams permit the pilot of an incoming plane to determine its horizontal, vertical, and longitudinal orientation during its approach. Earlier types of equipment were also described.

Earl Kiernan, maintenance engineer for Western Air Express, pre-

sented a paper on "Aviation Radio Equipment" in which he described standard equipment employed by his company.

Hugh Duckworth, assistant manager of the airways Traffic Control Division of the Department of Commerce, discussed the "Uses of Radio in Airways Traffic Control." This covered regulations governing both established airlines and the itinerant flier and the method of maintaining this regulation by radio.

Herbert Hela, operator in charge for the Los Angeles area of the Department of Commerce, spoke on the "Operation of Radio Facilities." This covered the various types of stations and the kinds of traffic handled by them.

Joseph Gurr, United Airlines maintenance engineer, presented "Aviation Radio Developments since 1928." He described briefly the problems of weight, frequency maintenance, and static.

The symposium was closed with a paper by R. O. Brooke of the National Broadcasting Company who discussed "Airplane Broadcast Pickups" which covered not only the actual operation but the equipment used in it.

Montreal

The Montreal Section met in the Engineering Institute of Canada auditorium on March 23. A. M. Patience, chairman, presided and there were thirty-eight present.

J. P. Henderson of the Dominion Observatory at Ottawa, presented a paper on "Radio, Sunspots, and Observatory Time." It was pointed out that sunspots have certain internal features, since they are surface storms or whirling upheavals of matter at temperatures as high as 6000 degrees centigrade. This is about twice the highest temperature that has been produced on the earth. Sunspots appear in 11.2-year cycles and in regular locations on the surface of the sun. They rotate at its 26-day period but not at all regions together as would be expected were the sun solid. The long spot periods have certain coincidences with the earth's climate and electric- and magnetic-storm frequency which indicates the probability of a causative relation. Researches in high-frequency transmission and of the recording of magnetic and auroral storms is being attempted to disclose such correlation as may exist.

Time services of observatories include many which are not generally known. Submarine and subterranean prospecting, location of oil fields and submarine-cable faults, dangers to dam and bridge structures, synchronizing recording seismographs and other recording clocks located at distances from each other, and checking of frequency stand-

ards are among them. In the group of well-known uses are oceanic navigation, surveying electrical systems, light and radio beacons, fog warnings, radio operations, and others.

Time subdivision is accomplished by pendulums, tuning forks, and piezoelectric-crystal oscillators. Comparisons can be made by a vernier with an electrical contact which can be set at any fraction of a second on the shaft of a 1000-cycle synchronous motor. Other methods employ the oscilloscope and the chronograph, which employs a rotating drum which makes one revolution or 24 inches of pen travel each second. A gaseous tube may be flashed to illuminate the face of a high-speed clock pointer and the time of the flash read by eye or photographed.

New York

The April 6 New York meeting was held in the Engineering Societies Building with H. M. Turner, a member of the Board of Directors, presiding in the absence of President Pratt.

A paper on "The Fine Structure of Television Images," by H. A. Wheeler and A. V. Loughren of the Hazeltine Service Corporation was presented by Mr. Wheeler. The paper appears elsewhere in this issue.

The attendance totaled four hundred and fifty.

Philadelphia

On March 3, the Philadelphia Section met at the Engineers Club. Irving Wolff presided in the absence of the chairman and there were two hundred present.

"The Nature and Properties of Wave-Guide Transmission" was the subject of a paper by A. E. Bowen of the Bell Telephone Laboratories. Experiments were performed to show how ultra-high-frequency waves may be guided through pipes filled with air or other dielectrics and along dielectric rods. Using a small antenna and crystal detector, the electric and magnetic lines of force in four types of waves were outlined. It was shown that the diameter of the pipe must exceed a certain critical value before the waves can be transmitted through it. Although most of the demonstration was with short pieces of pipe, the use of high-dielectric-constant material and an increase in frequency permitted the utilization of smaller diameter pipes. A 3000-megacycle wave was transmitted through a short length of 1-inch diameter pipe containing a rubber dielectric and through fifty feet of flexible pipe $2\frac{1}{4}$ inches in diameter. A receiver designed to terminate the long pipe in its characteristic impedance was demonstrated. The paper was concluded with some experiments on a resonant cavity used as a wave meter and as a resonant transformer for impedance matching.

Pittsburgh

The March 8 meeting of the Pittsburgh Section was held jointly with the Engineers Society of Western Pennsylvania and the local section of the American Institute of Electrical Engineers, at the William Penn Hotel. It was presided over by Vice-Chairman Furst of the electrical engineering group and attended by two hundred.

A paper on "Interesting Features Found in Indicating Instruments," was the subject of a paper by H. L. Olesen, general sales manager of the Weston Electrical Instrument Corporation. There was presented first a detailed description of the various parts from which the moving elements of indicating instruments are constructed. Samples of these parts were available for inspection. The second part of the paper covered various indicating instruments and their uses.

On March 14 the Pittsburgh Section met at Mellon Institute auditorium with R. T. Gabler, chairman, presiding. There were one hundred and twenty present.

"A Demonstration of Propagation of Ultra-High-Frequency Radio Waves Through Guides," was the subject of a paper by G. C. Southworth of Bell Telephone Laboratories. This was the same material as presented at the February 2 New York meeting of the Institute which is reported on page 261 of the March, 1938, Proceedings.

San Francisco

On March 2, a seminar meeting of the San Francisco Section was held at Manning's Coffee Cafe, with C. J. Penther, vice chairman, presiding. There were fourteen present.

Three papers were discussed. The first on "Frequency Discrimination by Inverse Feedback," by A. H. Fritzinger was published in the February, 1938, issue of the Proceedings and reviewed by Clark Cahill. The second paper "A New Type of Selective Circuit and Some Applications," by H. H. Scott, which also appeared in the February, 1938, issue, was reviewed by E. Ginzton. H. Thal-Larsen led the review on "The Basic Principles of Superregenerative Reception," by F. W. Frink, which appeared in the January, 1938, issue.

The March 16 meeting was attended by thirty-two and held in the Pacific Telephone and Telegraph Company auditorium. Noel Eldred, chairman, presided.

R. C. Shermund, engineer for Heintz and Kaufman, Ltd., presented a paper on "The Production of Metal Receiving Tubes." Details of the construction and assembly of the various elements used in metal tubes were given. The many difficulties encountered in their manufacture were pointed out and methods of overcoming them described.

Seattle

A. R. Taylor, chairman, presided at the March 25 meeting of the Seattle Section held at the University of Washington and attended by forty-eight.

"Volume-Level Problems in Radio Broadcasting" was the subject of a paper by M. V. Kiebert, inspector in the Seattle office of the Federal Communications Commission. The economical and engineering limitations on the transmitted dynamic range used in broadcasting were discussed. It was pointed out that for high fidelity, compression at the transmitter and expansion at the receiver must be reciprocally related. However, some compression, either manual or automatic is necessary to keep program levels between the limits of room noise and the permissible peak volume at the receiving point. It was suggested that the optimum dynamic range for broadcasting was between 40 and 55 decibels. Continuous compression systems and limiting systems were discussed. Various circuits used in commercial application of these types were described. The time constants for attack and decay were considered on the basis of the characteristics of the ear. It was suggested that standards might be adopted for compressors which would permit the use of expanding devices for receivers which would improve reception but not make reception without the expanding device unsatisfactory.

In the discussion which followed, and was participated in by Messrs. Libby, Walker and others, present commercial trends toward additional compression of recorded and wire-relayed programs was covered.

Toronto

Two meetings of the Toronto Section were held during March at the University of Toronto and presided over by W. H. Kohl, chairman.

The meeting on the 14th was attended by sixty-five and a paper on "Nuclear Physics," was presented by Joseph Thwaites of the Canadian Westinghouse Company, Ltd. He described in some detail the history of nuclear physics giving definitions of the terms atoms, molecules, electrons, and nucleus. Their relations to each other were shown. The difficulties of nuclear disintegration and the energies required to produce such phenomena were discussed. The cyclotron and its operation was described as was the high-voltage static machine built by the Westinghouse organization. A description was given of the Wilson cloud chamber and the theory of its operation. The results of bombarding various elements with neutrons were described and it was pointed out that the efficiency of the apparatus was vanishingly low. While transmutation from one element to another could be effected, the proc-

ess at present is very costly. The paper was discussed by Messrs. Hepburn, Kohl, Leslie and others.

There were one hundred present at the March 28 meeting at which A. S. Blatterman, design engineer and plant manager of the Rogers Majestic Corporation presented a paper on "The Design of Radio Receivers." Mr. Blatterman outlined the various factors which influence receiver design. The first was the geography of the market in which the set is to be used. Densely populated areas permit the sale of a large number of sets having relatively poor sensitivity as contrasted with the conditions in less densely populated areas located remotely from transmitting stations. Next to be considered is the wealth of the market. The sets must be designed to meet existing economic conditions. The third factor concerns the fashions of the market to be reached and includes types of cabinets desired, dials, and other factors which influence the appearance of the set. The manufacturer's capabilities must be considered and include financial as well as production possibilities. The fifth consideration was profit, which is based on manufacturing costs, as well as sales expense, and overhead. In designing a receiver the sales department specifies what is wanted and what the selling price must be. Preliminary work is done on the design of the chassis and costs computed. The final design is a balance to meet economically the requirements of the sales department. A number of features such as noise suppression, automatic tuning, and methods of compensating for inaccurate tuning were discussed from a designer's viewpoint. The use of broad-nosed steep-sided selectivity characteristics was treated.

Washington

On March 14 the Washington Section met in the Potomac Electric Power Company auditorium. E. H. Rietzke, chairman, presided and there were one hundred and thirty-five present.

"Some Problems of Broadcast Engineering" was the subject of a paper by R. F. Guy, radio facilities engineer, National Broadcasting Company. He discussed the problems of broadcast transmission coverage, the reduction of fading, and other related factors. The problems of antenna design particularly where limitation of height is necessary, were considered and methods described by which antennas considerably less than optimum height were made to function nearly as satisfactorily as the taller structures. The paper was discussed by Messrs. Davis, Rietzke and others.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

TECHNICAL PAPERS

HERTZ, THE DISCOVERER OF ELECTRIC WAVES*

By
Julian Blanchard
(Bell Telephone Laboratories, Inc., New York, N.Y.)

FIFTY years have passed since those memorable researches of the young German physicist, Heinrich Rudolph Hertz, which have come to be regarded as the starting point of radio. For it was he who first detected, and measured, electromagnetic waves in space—waves which had been predicted, it is true, but which had never before been observed. It is not to be claimed, of course, that the radio art would have failed to be born were it not for his genius, for we know that almost simultaneously the experiments of Lodge in England were pointing with certainty to the same discoveries, and the speculations of others were revolving around the possibility of generating electric waves. Yet it was the remarkably clear vision of Hertz, combined with his consummate persistence and skill, that won for him the prize and justly enshrined his name among the immortal men of science.

So, upon this golden anniversay of the opening of a new epoch in the realm of communication, it is fitting that we pause to do honor to his memory and to consider anew the significance of his great accomplishment.

The formal facts of Hertz's biography can be set down very quickly. He was born at Hamburg on February 22, 1857, his father an attorney, belonging to a family of successful merchants, his mother the daughter of a doctor of medicine, and the descendant of a long line of Lutheran ministers—all of cultural tastes and attainments on both sides. At the age of twenty he went off to school at Munich, after a rather unorthodox preparatory training, supposedly to pursue an engineering career, but he was torn between this resolve and his natural inclination for the study of pure science. Soon after reaching Munich he felt compelled to put the matter before his parents, to whom he frequently and confidingly wrote concerning his plans and his work. In a long letter written in November, 1877, he said, "I really feel ashamed to say it, but I must: now at the last moment I want

* Decimal classification: R097. Original manuscript received by the Institute, February 8, 1938.

to change all my plans and return to the study of natural science. I feel that the time has come for me to decide either to devote myself to this entirely or else to say good-bye to it; for if I give up too much time to science in the future it will end in neglecting my professional studies and becoming a second-rate engineer. Only recently, in arranging my plan of studies, have I clearly seen this—so clearly that I can no longer feel any doubt about it . . ." And then follows a lofty and appealing presentation of the reasons for his choice. Concluding, he wrote: "And so I ask you, dear father, for your decision rather than for your advice; for it isn't advice that I need, and there is scarcely time for it now. If you will allow me to study natural science I shall take it as a great kindness on your part, and whatever diligence and love can do in the matter, that they shall do. I believe this will be your decision, for you have never put a stone in my path, and I think you have often looked with pleasure on my scientific studies . . ."

Matters were arranged as he wished and he joyfully pursued his studies at the University and at the Technical Institute, working hard on mathematics and mechanics, and spending much time in the physical laboratory. In October, 1878, he went to Berlin and became there a student under the mighty giants, von Helmholtz and Kirchhoff; writing to his parents in November, "I am now thoroughly happy, and could not wish things better." In 1880 he gained his doctorate, and in October of that year was appointed assistant to Helmholtz, which delightful and stimulating association continued until Easter of 1883. He then went to the University of Kiel to become lecturer in theoretical physics, and here he first began to reflect seriously upon Maxwell's electromagnetic theory of light.

He was soon promoted again, and at Easter, 1885, he became professor of experimental physics at the Technical High School in Karlsruhe. Here, in 1886, at the age of twenty-nine, he married Elizabeth Doll, daughter of the professor of geodesy in the same institution, and their home became a congenial meeting place for their many cultured associates. It was at Karlsruhe also that he began his researches on electric waves. Before they were finished he was called in 1889 to succeed the celebrated Clausius at the University of Bonn, thus at the age of thirty-two having arrived at a position in the academic world not ordinarily to be achieved until much later in life.

He soon thereafter relinquished to others the further exploration of the great new territory of electric waves he had opened up and returned to some investigations on the discharge of electricity in rarefied gases, a subject which had interested him while at Berlin. He then devoted his attention entirely to what proved to be his last work, a

treatise on mechanics. In the summer of 1892 he suffered a severe illness which eventually led to chronic blood poisoning, of which he died, after indescribable suffering, on January 1, 1894. He would have been just thirty-seven years old the following month.

Although the fame of Hertz rests primarily on his electric-wave researches, these constitute by no means the whole of his work. His collected papers, edited by the German physicist, Dr. Philipp Lenard, and admirably translated into English by Professor D. E. Jones and associates, comprise three volumes. The first consists mainly of his miscellaneous earlier papers, some twenty-odd titles altogether. One of these, published in 1884, "On the Relations between Maxwell's Fundamental Electromagnetic Equations and the Fundamental Equations of the Opposing Electromagnetics," marked an important step in the development of Hertz's ideas, and has been called his greatest contribution to theoretical physics. In it he opposed the old orthodox theories of electric phenomena based on action at a distance, which were supported by most of the Continental physicists, and definitely aligned himself with the followers of Maxwell. This volume also contains his semipopular Heidelberg lecture of 1889, "On the Relations between Light and Electricity," giving a general account of his more recent work. It strikingly illustrates the charm and felicity of style he could employ in the presentation of a difficult subject, and cannot fail to be read, and reread, with pleasure and admiration. The volume ends with a eulogy of Helmholtz on the occasion of his seventieth birthday, wherein the pupil, in equally graceful language, paid homage to his beloved and inspiring preceptor. These two papers reveal so clearly between the lines the manner of man their author was.

The second volume contains the papers on electric waves, which had been collected by the author himself, as a result of numerous requests for reprints, and published under the German title, "Untersuchungen über die Ausbreitung der Elektrischen Kraft," and later in English as "Electric Waves," with a lucid introduction by Hertz explaining the motive and significance of each of the separate papers. One of the first of these describes an important by-product of the electric-wave studies, the discovery of the effect of ultraviolet light upon an electrical discharge. This discovery itself uncovered a new wealth of physical problems and the subject became immediately of great interest to many experimenters, most of whom at the time little suspected that this subsidiary effect was not the main discovery, or imagined that electrical science was on the eve of much greater conquests. The paper attracted attention to Hertz and aroused a popular interest in him, so that everything coming from him thereafter was

eagerly read. Had it not been for this rather accidental circumstance the importance of his subsequent work might have gone temporarily unnoticed, as has happened with some of the greatest discoveries.

The third volume in the series is his book, "Die Prinzipien der Mechanik," completed with difficulty during his last illness and published a few months after his death. A lengthy preface by the venerable Helmholtz gives an appreciative sketch of the author's life and work —this time a sorrowful tribute from master to pupil.

In order to understand and appraise the work of Hertz on electric waves, it will be needful to review briefly the ideas about light and electricity that prevailed at his time and before. With regard to light, Newton's corpuscular theory had given way in the early 1800's to the wave theory of Young and Fresnel, and long before Hertz's day the idea of transverse vibrations in a hypothetical elastic-solid medium called the luminiferous ether had become firmly established. Length of wave and velocity had been measured many times. The wave theory accounted satisfactorily for all the known phenomena in optics and there was no doubt in anybody's mind about its essential correctness, regardless of any difficulties encountered in explaining the nature of the ether and its relation to matter.

As to electricity and magnetism, the older theories of instantaneous action at a distance were beginning to weaken with the discoveries, in the early part of the nineteenth century, of the reactions between electric currents and magnets, and the phenomenon of induction. Hitherto there had been no postulation of an intervening medium to explain the transmission of the force between two charged bodies, and it was supposed that electricity and magnetism, like gravitation, acted across empty space in straight lines and instantaneously. In somewhat different dress these ideas were given new life around the middle of the century, particularly by some of the German physicists. To Faraday, however, such views were unacceptable. He wished to get rid of the idea of action at a distance and in his mind pictured a medium, along the contiguous molecules or particles of which the force was propagated. In this medium he visualized "lines of force" emanating from or terminating upon the electric charges or magnetic poles, acting like stretched elastic threads, repelling each other sidewise as well as tending to contract. Thus, in his thinking, attention was focused upon the insulating medium surrounding a conductor, the "dielectric" as he called it, for here, he thought, was the real seat of the action. He believed also that there existed some direct relation between light and electricity or magnetism. He was ever seeking to find such a relation and in the course of his many experiments he discovered the rotation

of the plane of polarization of light by a magnetic field. In his speculations there was one question which was continually presenting itself to him. Do electric and magnetic forces, like light, require time for their propagation? Are there waves? But to this question he was unable to find an answer.

And then came Maxwell, building upon the foundation that Faraday had laid, translating Faraday's ideas into the language of mathematics, and making the grand generalization that light and electric waves are one and the same phenomenon, propagated by the same medium, with the same velocity, and differing only in wavelength. Like Faraday, he considered the energy of the electromagnetic field to reside in the dielectric. He conceived the medium to have properties analogous to those of an elastic solid, which would spring back to its original state upon the removal of the straining force. The alteration of the displacement, or "polarization," in the medium was viewed by him as an electric current, which he called the "displacement current," as distinguished from the "conduction current" existing in conductors. From the general equations which he formulated it was shown that only transverse vibrations (like light) could be propagated in such a medium and that the velocity of propagation should be equal to the ratio of the two systems of electrical units, that is, to the number of electrostatic units of electricity contained in one electromagnetic unit. This ratio had been experimentally determined by Weber and Kohlrausch (it was later redetermined by Maxwell himself, by a different method), and the fact that it agreed so very closely with the measured velocity of light was one of the strongest points in favor of the view that light waves were identical with the hypothetical electromagnetic waves. Maxwell's comprehensive theory was first enunciated in his 1865 paper entitled "A Dynamical Theory of the Electromagnetic Field," and afterwards elaborated in his great "Treatise on Electricity and Magnetism," published in 1873, but for a number of years it was regarded by many as merely a speculation, by others as probably true, and by none as conclusively proved. It remained for Hertz to add the capstone to the theory by actually demonstrating for the first time the existence of electromagnetic waves in space; and furthermore, to show experimentally that they had all the physical properties of ordinary light waves.

In 1888, while he was teaching at the Technical High School in Karlsruhe, Hertz carried out the brilliant experiments which have made his name famous. These were actually a part of a long series of experiments which began in 1886, and which came about partly by accident, and yet were the result of his keen interest in everything

connected with electric oscillations; an interest extending back to 1879, when, at the suggestion of Helmholtz, he had considered tackling a prize problem proposed by the Berlin Academy of Science aimed at the proof of a portion of Maxwell's theory, but which he had abandoned for the reason that oscillations of sufficiently high frequency were not then available. While using in his lectures at Karlsruhe a pair of flat ("pancake") coils, called Reiss or Knockenhauer spirals, mounted adjacent to each other, he was surprised to find how easy it was to obtain sparks between the terminals of the secondary coil when a small Leyden jar or even a small induction coil was discharged through the primary, provided the primary discharge took place across a spark gap. This, of course, was an indication of an exceptionally strong inductive effect. This observation led to his discovery of a method of exciting electric disturbances far more rapid than any hitherto known, such as those of Leyden jars or open induction coils as customarily used, and having wavelengths, it turned out, capable of being measured within the confines of a laboratory. His oscillator was nothing more than a short metal rod with a spark gap in the middle (sometimes with metal spheres or plates attached to the ends, resembling a dumbbell), the sparking terminals consisting of small knobs or spheres which were connected to the terminals of a Ruhmkorff induction coil; the small inductance and capacitance of this simple linear conductor, together with the proper functioning of the spark gap, accounting for its success. By such means Hertz obtained wavelengths from a few meters down to 30 centimeters, and so began, it is seen, with the "ultrashort" waves that are again coming into vogue.

Hertz began his experiments with a study of the "induction" about this exciter. As he commented in his first paper in this series, "On Very Rapid Oscillations," published in May, 1887, theory had predicted the possibility of very rapid oscillations in open-wire circuits of small capacitance, but it could not be predicted from theory whether they could be produced on such a scale as to admit of their being observed. Hertz not only devised a method of producing such oscillations, but also discovered a method of detecting them, by their effects in the surrounding space. His detector consisted merely of a short length of wire bent in the form of a rectangle or a circle and containing a micrometer spark gap, across which minute sparks could be seen in a darkened room; especially if this secondary circuit was in electrical resonance with the exciter. This exceedingly simple detector was indeed a capital discovery. Some five years earlier Professor G. F. Fitzgerald, of Dublin, had suggested "the combination of a vibrating generating circuit with a resonant receiving circuit . . . as one by which this very

question might be studied." But, as he said afterwards in speaking of Hertz's work, "I did not see any feasible way of detecting the induced resonance: I did not anticipate that it could produce sparks." Concerning this contrivance the following interesting remarks were made by its author in his Heidelberg lecture above referred to: "The method had to be found by experience, for no amount of thought could well have enabled one to predict that it would work satisfactorily. For the sparks are microscopically short, scarcely a hundredth of a millimeter long; they only last about a millionth of a second. It seems absurd and almost impossible that they should be visible; but in a perfectly darkened room they *are* visible to an eye which has been rested in the dark. Upon this thin thread hangs the success of our undertaking." *Multum in parvo*, truly!

After a series of preliminary experiments, in which he studied the various induction effects, including the phenomenon of resonance, and demonstrated waves on wires (an earlier, but overlooked, discovery of von Bezold, in 1870), and also solved the problem of the Berlin Academy—"to establish experimentally any relation between electromagnetic forces and the dielectric polarization of insulators"—he was fully convinced that the disturbance was propagated through space, independently of wires, with a finite velocity and in the form of waves, in accordance with Maxwell's prediction. His conclusion was then definitely and convincingly proved by making use of the well-known method of reflection and interference to produce standing waves, and noting the position of the nodes and antinodes. These epochal experiments were described in a paper entitled "Electromagnetic Waves in Air and Their Reflection," published in May, 1888. But he did not stop there, and in these and succeeding investigations he showed that electric waves are reflected from plane and curved metal surfaces in accordance with the same laws as light waves; that they are refracted in passing through prisms of pitch, paraffin, and other dielectrics; and that they are polarized by a grating of parallel wires, and hence are transverse waves. From actual measurements of their wavelength and computations of their frequency (from the constants of his oscillator), he calculated their velocity and found that it was the same as the velocity of light. As summarized by Hertz himself, "The object of these experiments was to test the fundamental hypotheses of the Faraday-Maxwell theory, and the result of the experiments is to confirm the fundamental hypotheses of the theory." The old action-at-a-distance philosophy had come to an end.

The importance of Hertz's contributions to this great subject received instant and enthusiastic recognition, and his experiments were

soon being repeated in all the important laboratories of the world. The English mathematical physicist, Oliver Heaviside, writing in 1891, said: "Three years ago electromagnetic waves were nowhere. Shortly after, they were everywhere." Here were researches of a most abstruse and complex character, with no apparent utility and having no elements of popular appeal, and yet bringing to their author such acclaim as had seldom been accorded to a man of science. Honors were showered upon him on every hand, at home and abroad. In England, where his work was especially appreciated, he was awarded the coveted Rumford medal by the Royal Society.

Hertz's characteristic modesty in referring to his own achievements was matched only by his generosity in giving credit to the accomplishments of others. In one of his lectures he said, "Such researches as I have made upon this subject form but a link in a long chain . . . Lack of time compels me, against my will, to pass by the researches made by many other investigators; so that I am not able to show you in how many ways the path was prepared for my experiments, and how near several investigators came to performing these experiments themselves." Mention has been made of the investigations of Sir Oliver Lodge in the same field and the imminence of his discovery of the same phenomena. It is pleasant, indeed, in this instance to be able to record the absolute lack of any feeling of jealousy or envy on the part of either of these courteous gentlemen. In the introduction to his collected papers Hertz wrote, "I may here be permitted to record the good work done by two English colleagues who at the same time as myself were striving towards the same end. In the same year in which I carried out the above research, Professor Oliver Lodge, in Liverpool, investigated the theory of the lightning conductor, and in connection with this carried out a series of experiments on the discharge of small condensers which led him on to the observation of oscillations and waves in wires. Inasmuch as he entirely accepted Maxwell's views and eagerly strove to verify them, there can scarcely be any doubt that if I had not anticipated him he would have succeeded in observing waves in air, and thus also in proving the propagation with time of electric force. Professor Fitzgerald, in Dublin, had some years before endeavored to predict, with the aid of theory, the possibility of such waves, and to discover the conditions for producing them."

On his part Lodge just as generously wrote, only a few years afterwards, in an obituary of his rival: "Hertz stepped in before the English physicists, and brilliantly carried off the prize. He was naturally and unaffectedly pleased with the reception of his discovery in England, and his speech on the occasion of the bestowal of the Rumford medal

by the Royal Society will long be remembered by those who heard it for its simplehearted enthusiasm and good-feeling. His letters are full of the same sentiment . . ."

Noteworthy, indeed, was the extreme modesty of this scientific lion of the hour, and equally striking his consideration for the feelings of others. It is recorded that when the Royal Society presented him with the Rumford medal, "he silently disappeared from Bonn for a few days—none knew why—and he returned as silently." In refusing the request made by the editor of *The Electrician* (of London) in 1890 for his photograph, Hertz replied, "I feel as if presenting my portrait now in so prominent a place follows too quickly the little work I have done. I should like to wait a little, and see if the general approbation which my work meets with is of a lasting kind. Too much honor certainly does me harm in the eyes of reasonable men, as I have sometimes occasion to observe. If your kind intention is the same in two years, even one year, I shall readily consent and help you in every respect." Four years later the portrait was published, following upon Hertz's death.

Upon the untimely ending of his brief but brilliant career, occurring in the very prime of life, before he was yet thirty-seven, there was a feeling of shock and sadness in every scientific quarter. Many were the sincere tributes paid to his memory, honoring him for his rare personal qualities as well as his distinguished scientific attainments. Some expressions from Lodge have already been quoted. Said he in his obituary in *The Electrician*, "Not a student of physical science on the planet but will realize and lament the sad loss conveyed by the message, 'Hertz is dead'." The editor of that journal wrote, "In the modesty and self-forgetfulness which blend so admirably in the spirit of true scientific research Hertz was singularly rich." In an editorial note in an American journal, *The Physical Review*, we find the following: "In addition to the recognition of those who were able to appreciate his work, Hertz received the acclamations of the entire world of thought. Fortunately he possessed a nature of such complete simple-mindedness that his sudden rise into a position akin to notoriety had no effect upon him. The unassuming bearing which had always characterized him remained with him to the end."

In a memorial address delivered by Professor Herman Ebert before the Physical Society of Erlangen on March 7, 1894, the following sentiments are expressed: "In him there passed away not only a man of great learning, but also a noble man, who had the singular good fortune to find many admirers, but none to hate or envy him; those who came into personal contact with him were struck by his modesty and

charmed by his amiability. He was a true friend to his friends, a respected teacher to his students, who had begun to gather round him in somewhat large numbers, some of them coming from great distances; and to his family he was a loving husband and father."

It can be said in retrospect that the fundamental invention in radiotelegraphy was made by Hertz, and yet it is true that the discoverer of electric waves had no anticipations as to their utilitarian possibilities. There was no rush to the patent office; indeed, it was not until two years after Hertz's death that the first application for a radio patent was filed, by Marconi. The chief interest at the time was purely scientific, the results being hailed as the settlement of a great scientific controversy, the confirmation of Maxwell's theory, the annexation to electricity of the entire domain of light and radiant heat. In the current literature we find little of prophecy with respect to utility. Sir William Crookes has been credited with being one of the first to foresee distinctly the applicability of "Hertzian" waves to practical telegraphy. In an article in the *Fortnightly Review* for February, 1892, he made a remarkably accurate forecast of what was to come: "simpler and more certain means of generating electrical rays of any desired wavelength"; "more delicate receivers which will respond to wavelengths between certain defined limits and be silent to all others"; "means of darting the sheaf of rays in any desired direction . . ." And for secrecy he foresaw that "the rays could be concentrated with more or less exactness on the receiver," if the sender and receiver were stationary; or, if moving about, "the correspondents must attune their instruments to a definite wavelength . . ." "This is no mere dream of a visionary philosopher," he wrote. "All the requisites needed to bring it within the grasp of daily life are well within the possibilities of discovery, and are so reasonable and so clearly in the path of researches which are now being actively prosecuted in every capital of Europe that we may any day expect to hear that they have emerged from the realms of speculation into those of sober fact."

As we well know, all that he predicted, and more, has become reality, although progress was not to be as rapid as then seemed probable. One of those who at the time had the imagination to see, if only hazily perhaps, the great possibilities of Hertz's discovery was the youthful Marconi, who had also the initiative and the determination to put his ideas into execution, to make the new-found waves useful to mankind. Within a few years, around the turn of the new century, the world was to be thrilled by the detection of a wireless signal transmitted across the wide Atlantic. But there were insurmountable limitations to the means at hand, and it remained for still another wave in the

onward roll of science, the advent of the magical era of electronics, to yield the tools necessary for the really great advance that was ahead. With the invention and development of the amplifying and oscillating vacuum tube progress was greatly accelerated, and in a comparatively short time there had been achieved, by a veritable army of experimenters, the marvels of world-wide intercommunication which are so familiar to us today.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

STABILITY OF TWO-METER WAVES*

By
Charles R. Burrows, A. Decino, and Loyd E. Hunt
(Bell Telephone Laboratories, Inc., Dial, New Jersey)

Summary—*The continuous records of the field strength received over a 60-kilometer path on a frequency of 150 megacycles for the year 1936 are analyzed. Preliminary comparison with other paths of the same length indicate that the magnitude of the recorded variations of the signals may be typical of paths of this length.*

A reduction in the path length by a factor of two reduced the fading range in decibels by a factor of five.

The results are found to be in agreement with an earlier formula. Fading reduced the field 7 decibels below the average value 1 per cent of the time.

Introduction

EARLY measurements made on ultra-short-wave propagation gave the general impression that these signals were extremely stable in magnitude. In 1931 Jouaust[1] reported instability of ultra-short-wave signals. Since that time, numerous investigators[2,3,4,5,6,7,8] have noticed that ultra-short-wave signals were not stable when the attenuation in addition to the inverse distance attenuation was large, or when the receiver was below the line of sight. The present paper reports an intensive investigation of ultra-short-wave stability in which fading was observed even for conditions where the additional attenuation was small and also for optical-path transmission.

* Decimal classification: R113.7. Original manuscript received by the Institute November 26, 1937. Presented before the joint meeting of the U.R.S.I. and the I.R.E., Washington, D.C., April 30, 1937, and the meeting of the Radio Wave Propagation Committee of the C.C.I.R., London, England, November 23–25, 1937.

[1] R. Jouaust, "Some details relative to propagation of very short waves," Proc, I.R.E., vol. 19, pp. 479–488; March, (1931). See p. 485.

[2] J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, "Ultra-short-wave propagation," Proc. I.R.E., vol. 21, pp. 427–463; March, (1933), and *Bell Sys. Tech. Jour.*, vol. 12, pp. 125–161; April, (1933). See p. 135.

[3] L. F. Jones, "A study of the propagation of wavelengths between three and eight meters," Proc, I.R.E., vol. 21, pp. 349–386; March, (1933). See pp. 373, 376–378.

[4] B. Trevor and R. W. George, "Notes on propagation at a wavelength of seventy-three centimeters," Proc. I.R.E., vol. 23, pp. 461–469; May, (1935). See p. 463.

[5] C. R. Englund, A. B. Crawford, and W. W. Mumford, "Further studies of ultra-short-wave transmission phenomena," *Bell Sys. Tech. Jour.*, vol. 14, pp. 369–387; July, (1935). See pp. 376–378.

[6] Ross A. Hull, "Air-mass conditions and the bending of ultra-high-frequency waves," *QST*, vol. 19, pp. 13–18; June, (1935).

[7] C. R. Burrows, Alfred Decino, and Loyd E. Hunt, "Ultra-short-wave propagation over land," Proc. I.R.E., vol. 23, pp. 1507–1535; December, (1935). See pp. 1522–1524.

[8] Ross A. Hull, "Air-wave bending of ultra-high frequency waves," *QST*, vol. 21, pp. 16–18; May, (1937).

In an experimental investigation of ultra-short-wave propagation over land,[7] fading was observed on all experimental frequencies for the paths 44 and 72 kilometers long but not on the paths 9 and 26 kilometers long. This was taken as a confirmation of the earlier observation that fading occurs when the attenuation in addition to the inverse-distance attenuation reaches the order of 30 decibels. While these data are consistent with the present experimental data, we now know that more extended observations show that even when the additional attenuation is small, the signal may not be stable. The chances of finding this instability by 5-minute observations is small since the probability of fading as great as a decibel in any 5-minute interval is very small for the conditions of the experiment.

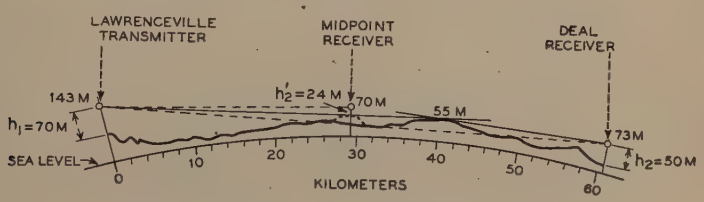


Fig. 1—Profile of Lawrenceville–Deal path.

Locations

The major part of the present data was obtained over a path between Lawrenceville and Deal, New Jersey. The choice of this path made it possible to take advantage of existing structures to support the transmitting and receiving antennas at 70 and 50 meters, respectively, above the ground. The profile of this 60.6-kilometer path is shown in Fig. 1. The length of this path is intermediate between those for which the earth's curvature may be neglected and those for which it has a pronounced effect. It is a nonoptical path as can be seen from the profile, but atmospheric refraction slightly greater than that which would increase the effective radius of the earth by a factor of four-thirds would make it optical.[9] The ratios of path length to shadow distances[7] without and with refraction are 1.26 and 0.79, respectively, so that the effect of the earth's curvature should just be noticeable for this path.

[9] The actual path length, 60.6 kilometers, is slightly longer than the longest optical path, 55.0 kilometers, that would result with the actual antenna heights if the irregularities of the earth's surface were neglected but is slightly shorter than the corresponding path, 63.6 kilometers, with a refraction factor of four-thirds.

Fig. 2—The transmitting antenna and supporting structure at Lawrenceville, New Jersey.

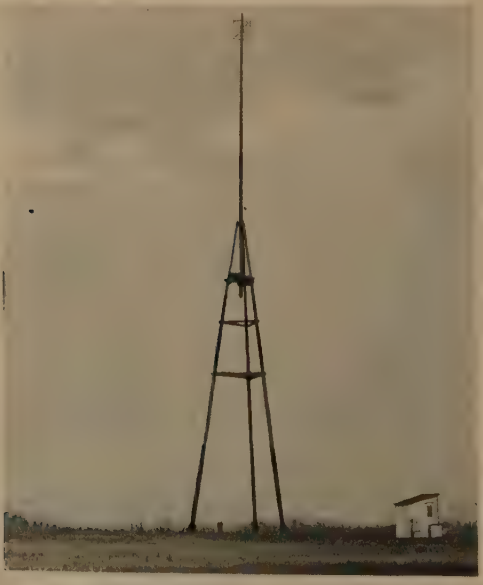


Fig. 3—The receiving terminal at Deal, New Jersey.

Equipment

The transmitting antenna at Lawrenceville, shown in Fig. 2, was supported by a 70-foot telephone pole thrust above a 200-foot steel tower. Any possible effect of the steel tower on the characteristics of the antenna was minimized by this separation and by directing a null downward. The receiving antenna at Deal, shown in Fig. 3, was supported by a 180-foot wooden structure.

The transmitting and receiving antennas consisted of 4 and 6 horizontal half-wave elements, respectively. Accordingly, they were slightly directive in both planes. For the major part of the test, both antennas were fed by coaxial transmission lines which were filled with

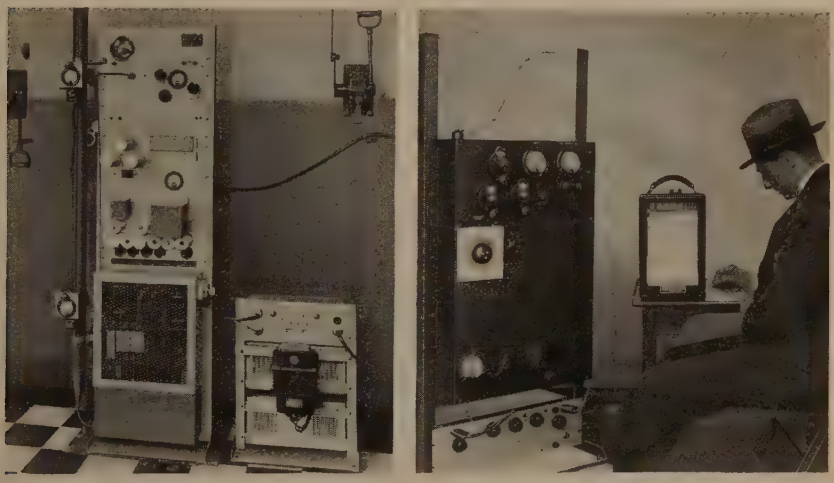


Fig. 4—Left, transmitting equipment. Right, receiving equipment.

nitrogen under pressure. During the first few months of the test a closely spaced open-wire transmission line was used at the receiver. With this line the over-all gain of the receiving equipment decreased 10 to 20 decibels with the onset of rain, presumably due to salt on the insulators. In a long rain this gain was partially restored, even while the line was wet, suggesting the possibility that the rain had washed the disturbing impurities from the insulators. To eliminate these variations, this line was replaced by a coaxial transmission line. Experience with the coaxial line at the transmitter indicated the necessity of keeping these lines filled with nitrogen under pressure in order to maintain their attenuation at a constant value.

The transmitter (Fig. 4) was crystal controlled for frequency stability. Variations in the radiated power were minimized by maintain-

ing the input power to the transmitter constant by means of a voltage regulator. The transmitter delivered about 4 watts of radio-frequency power to the transmission line feeding the antenna. The current in one element of the antenna was read every two hours.

The receiver (Fig. 4) was a double detection set with a wide-band intermediate-frequency amplifier. A high degree of gain stability was insured by employing negative feedback in the intermediate-frequency amplifier, thermostatic control of room temperature, and voltage regu-

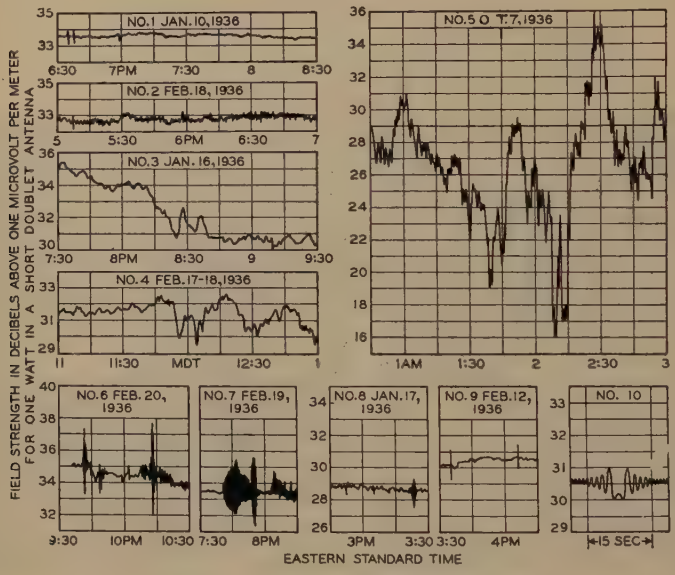

FIELD STRENGTH IN DECIBELS ABOVE ONE MICROVOLT PER METER FOR ONE WATT IN A SHORT DOUBLET ANTENNA
NO.1 JAN. 10, 1936
NO.2 FEB. 18, 1936
NO.3 JAN. 16, 1936
NO.4 FEB. 17-18, 1936
NO.5 O T.7, 1936
NO.6 FEB. 20, 1936
NO.7 FEB. 19, 1936
NO.8 JAN. 17, 1936
NO.9 FEB. 12, 1936
NO. 10
EASTERN STANDARD TIME
15 SEC

Fig. 5—Samples of fading records.

lation of the power supply. The signal was recorded graphically. The over-all gain of the receiving system was measured twice a day by a standard field-strength generator.

Fading Characteristics

The continuous recording of the received signal during the year disclosed several types of fading. Some samples are shown in Fig. 5. The first shows fairly steady reception with slow changes of a fraction of a decibel. Such steadiness often occurred for several consecutive days. In the second instance the variations have become rapid but were still small. No. 3 shows very slow fading of large amplitude with fading of a slow period superimposed. Most of the fading over this path was either slow or very slow as illustrated in Nos. 3 and 4. An interesting sample is that shown in 5, where slow fading was super-

imposed upon very slow fading with a range up to 20 decibels. During these periods there were occasions when the field dropped below the sensitivity of the receiver for a few seconds.

Some examples of rapid fading in which the period of oscillation was of the order of a second are shown in Nos. 6 to 10. Sometimes the oscillations lasted for about 15 minutes, while at other times they persisted for only about a minute. The variations indicated by the vertical lines of No. 9 are of particular interest in that each consists of 20 or 30 individual oscillations. One of these variations is shown with an expanded time scale in No. 10. A characteristic, common to all

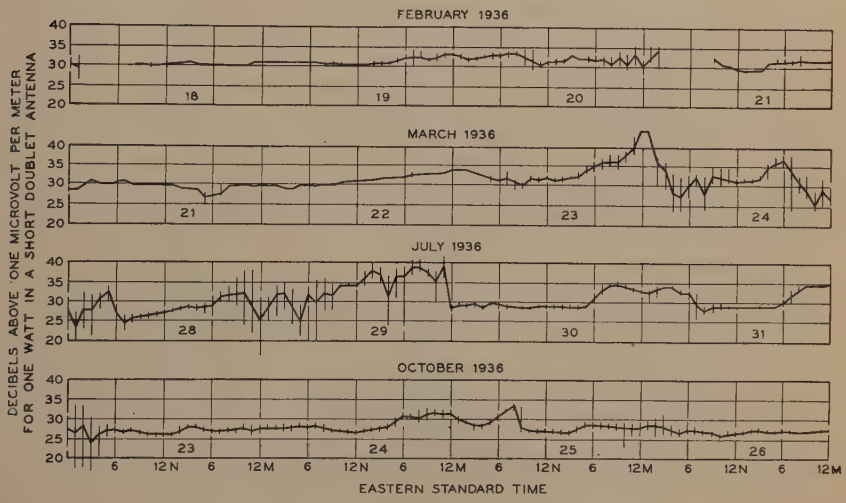


Fig. 6—Variation of signal over long periods.

of these patterns, is their symmetry about the center. This type of symmetry would result if the fading were caused by the interference of a wave reflected from some object, such as an air mass or an airplane, moving across the transmission path. Fading coincident with the passing of an airplane which would appear similar to this has been observed by many experimenters. For the year, both of these oscillatory patterns were more prevalent during the early spring and late fall and were seldom observed during the hottest summer months.

The variations of the signal over longer periods of time are summarized in Fig. 6. The continuous curve gives the average field strength for each hour. The vertical lines show the range of the signal within the hour. This figure shows that the field strength frequently remained constant to within a small fraction of a decibel for extended periods while at other times it varied in magnitude by as much as 15 decibels.

Diurnal Variations

An examination of the original records showed more pronounced fading and a tendency toward higher fields at night. Because of this diurnal variation in the fading magnitude, any diurnal variation of the average field strength would be dependent, to a certain extent, upon the method of computing the average field strength. The curves of Fig. 7 show the diurnal variation of the hourly average field strength. They represent the average of the field strength expressed in decibels, which is a geometric average of the field strength in microvolts per meter. While this is probably the most pertinent type of average from the point of view of the use of these signals as a means of communica-

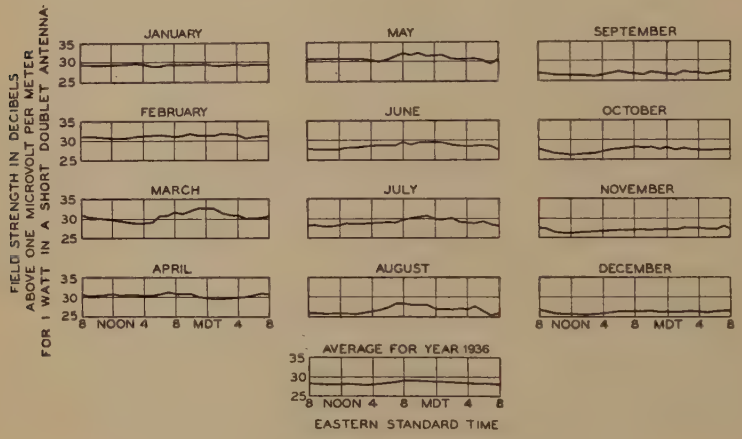


Fig. 7—Monthly average diurnal variation curves.

tion, it should be pointed out that an arithmetic average of the field strength expressed in microvolts per meter would show higher average fields for the hours of the day during which the field had the largest range of values. Even for the type of average used in Fig. 7 there is a definite tendency toward a maximum around 10 P.M. in several of the months. Although this maximum appears small in the yearly average it may represent something physically significant especially in view of the fact that an arithmetic average would show a greater maximum. In this connection it is interesting to note that Jouaust,[1] Trevor and George,[4] and Hull[6] also found this type of variation.

The diurnal variation of the fading range can be seen from Fig. 8 which shows the percentage of time that the hourly fading range was greater than any specified value as a function of the time of day. This shows that the field was steadiest around noon while most of the fading

occurred during the night. For instance, the hourly fading range around 1 P.M. was greater than 1 decibel only about 35 per cent of the time and it never exceeded 10 decibels, while near midnight the hourly fading range was greater than 1 decibel about 70 per cent of the time and exceeded 10 decibels about 5 per cent of the time.

Hourly Variations

An estimation of the stability of the signals over this path can be obtained from Fig. 9 which shows the percentage of time that the hourly fading range was less than any specified value for the entire

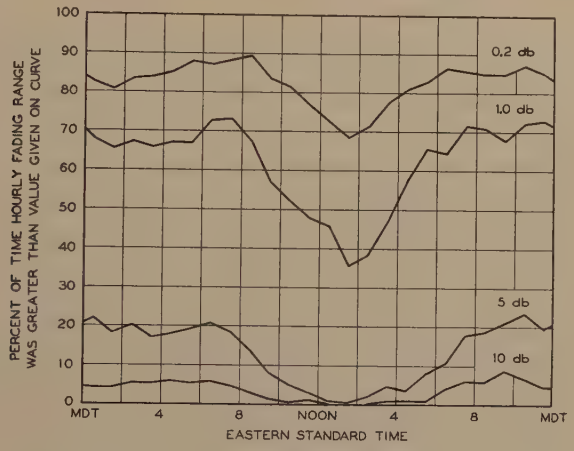


Fig. 8—Diurnal variation of fading magnitude.

year irrespective of the time of day. For example, the signals remained within a range of 1 decibel for a period of 1 hour for about 40 per cent of the time. Likewise the hourly fading range exceeded 16 decibels for only 1 per cent of the time.

Attenuation over the Path

As in previous work, the attenuation in addition to the inverse-distance attenuation, was measured by experimentally determining the free-space field radiated by the transmitting antenna.[7] The yearly average of this attenuation may be compared with our previously published theoretical formula. The ratio of useful received to radiated power between doublet antennas was given as[7]

$$\frac{P_R}{P_T} = \left[\left(\frac{3\lambda}{8\pi d}\right)\left(2 \sin \frac{2\pi h_1 h_2}{\lambda d}\right) \cdot F\right]^2$$

when either the argument of the sine function is less than unity or F

is substantially unity.[10] Here the three factors represent (1) the attenuation between short doublet antennas in free space, (2) the additional attenuation caused by negative reflection from the earth's surface assumed plane, and (3) the additional attenuation caused by the earth's curvature or the shadow factor,[11] respectively. For the conditions of the experiment these three factors expressed in decibels are $108.1+8.5+7.0=123.6$ decibels which corresponds to the measured value of 120.0 decibels for the yearly average. In view of the magnitude of the variation over this path, this agreement is satisfactory.

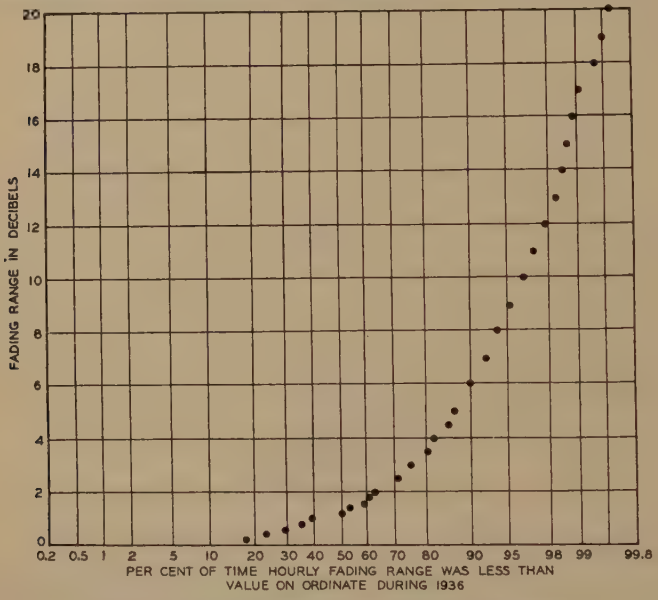


Fig. 9—Distribution of hourly fading ranges.

Van der Pol and Bremmer[12] have given a rigorous solution for transmission over a dielectric sphere and additional calculations along the same line have been made by M. C. Gray[13] for the paths used in this

[10] This formula is also limited to antenna heights for which the field is given by the straight part of curve 3 of Fig. 13 by Charles R. Burrows, "Radio propagation over plane earth—field-strength curves," *Bell Sys. Tech. Jour.*, vol. 16, pp. 45–75; January, 1937. In practice the antenna heights are usually within this region.

[11] In transcribing Fig. 10 of reference 7 for publication the curve was shifted to the right. To correct for this the labeling on the abscissa should be multiplied by 0.8. The asymptotic value of the shadow factor as stated in the summary of reference 7 is correct.

[12] Balth. van der Pol and H. Bremmer, "The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere, with applications to radiotelegraphy and the theory of the rainbow." *Phil. Mag.*, series 7, vol. 24, pp. 141–176, and 825–864; July and supplement to November, (1937).

[13] Bell Telephone Laboratories, Inc., New York, N.Y.

experiment. The following table compares the ratio of the received field with that which would result for transmission over plane earth under various conditions. While the experimental values were obtained

VALUES OF SHADOW FACTOR IN DECIBELS BELOW UNITY

$k=\frac{\text{effective radius of earth}}{\text{actual radius}}$	1	4/3
1. Early approximate theory[7]	8.8	7.0
2. Van der Pol and Bremmer's curve for vertical antennas on the ground[12]	15.4	11.8
Gray's calculations for antennas 70 and 50 meters high and		
3. $\epsilon = 4$	24.0	23.0
4. $\epsilon = 10$	20.7	19.9
5. $\epsilon = 30$	15.8	15.0
6. Average experimental value with horizontal antennas	3.4	

with horizontal antennas and the rigorous calculations are for vertical antennas we may tentatively apply them to the experimental data since available experimental information indicates no large difference between the average values of the attenuation with vertical and horizontal antennas under comparable conditions.

The latter values of the attenuation due to the shadow factor, which are based upon a rigorous mathematical analysis, are all considerably greater than the experimental value. The large difference between this experimental value and the theoretical values of the shadow factor indicates that idealized conditions upon which the theory is based, namely, spherical earth and homogeneous atmosphere, do not sufficiently approximate the experimental conditions. Also a uniform refraction of the order of magnitude that is likely to exist does not appreciably improve the agreement. Hence the assumption of a homogeneous atmosphere and ideally spherical earth is not justified even for paths as short as this.

The closer agreement of experiment with our shadow factor than with that of van der Pol and Bremmer has no theoretical significance. The difference between the two formulas represents the error introduced by the approximations made in our derivations which were pointed out at the time. In spite of this, our formula is in agreement with the mean values of experimentally determined fields in so far as applicable data are available. Hence our formula still provides a useful empirical indication of the average field.

While the average field strength may be calculated by the above formula, this alone does not allow the determination of the required transmitter power for a commercial ultra-short-wave circuit. Fading reduces the field strength below the average so that in order to maintain the required field-strength level for a certain percentage of the time, additional power must be provided. The additional power that

would have been required over this path during 1936 is shown in Fig. 10 as a function of the percentage of time the field may go below the specified value.[14] For example the field was more than 7 decibels below its mean value 1 per cent of the time. Hence after engineering a circuit on the basis of average field, an additional 7 decibels would be required in transmitter power in order to have sufficient field 99 per cent of the time.

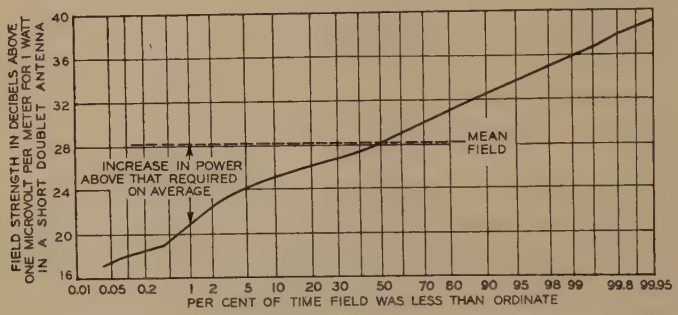

FIELD STRENGTH IN DECIBELS ABOVE ONE MICROVOLT PER METER FOR 1 WATT IN A SHORT DOUBLET ANTENNA
40
36
32
28
24
20
16
MEAN FIELD
INCREASE IN POWER ABOVE THAT REQUIRED ON AVERAGE
0.01 0.05 0.2 1 2 5 10 20 30 50 70 80 90 95 98 99 99.8 99.95
PER CENT OF TIME FIELD WAS LESS THAN ORDINATE

Fig. 10—Distribution of field strength.

Comparison of Paths

In order to determine whether the stability of ultra-short-wave propagation is better over a shorter path some measurements were made simultaneously over this path and one half as long. The profiles of these paths are shown in Fig. 1. Samples of these records are shown in Fig. 11. The receiving antenna height for the shorter path was about half that for the longer path so that the attenuation attributable to negative reflection was about the same for the two paths. On the average the fading range, measured on a decibel scale, was about one fifth as great over the shorter path as over the longer path. This indicates that the fading can be materially reduced by shortening the path length.

A few measurements were made over the longer path with lower antenna heights in an effort to determine the effect of antenna height on stability. During these limited tests, no difference in the fading was observed with a two-to-one change in antenna height. Besides the difference in height the two receiving antennas were 730 meters apart in the direction of transmission.

[14] For hours during which the average field was less than 25 decibels, the actual time the instantaneous field was below the specified value was employed in obtaining the curve. For the remainder of the curve the hourly average field was used instead of the instantaneous field to save labor. The two methods give the same result for fields of 23, 24, and 25 decibels and results differing by only a decibel for lower fields.

Some measurements were also made at Holmdel, New Jersey, simultaneously with measurements made by Englund, Crawford, and Mumford on Beer's Hill, of the field from their 2-meter transmitter at Lebanon, New Jersey. The variations of the signal at these two loca-

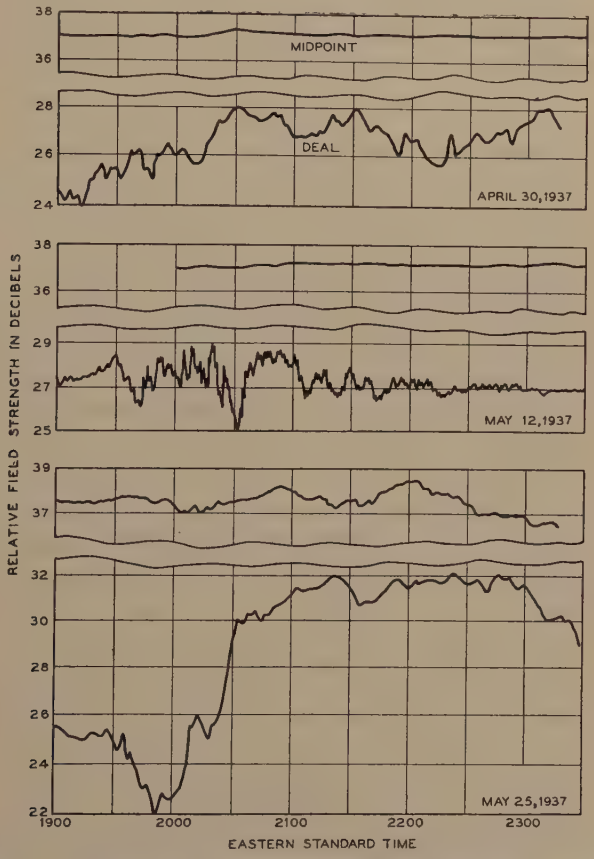


Fig. 11—Comparison of variations of field strength recorded at the mid-point and at Deal.

tions were very similar and most of the time practically identical although the receiving antennas were separated by 3.6 kilometers horizontally and 52 meters vertically. The absolute magnitude of the field strength was about 15 decibels less at the Holmdel location than at the Beer's Hill one. The larger, longer period variations were simultaneous and had the same amplitude and general shape. Sometimes the smaller, more rapid variations that were superimposed on the larger variations were not identical at the two locations.

The Lebanon-Beer's Hill path is about the same length as the Lawrenceville-Deal path but by locating the antennas on hills they are effectively 220 and 90 meters above the intervening ground. This makes the former path definitely an optical path, in fact, the received field strength is approximately the free-space field. In spite of these large differences in type of path, a preliminary comparison indicates the fading magnitude to be approximately the same over the two paths.

Conclusions

By continuously recording a signal of 150 megacycles over a 60-kilometer path for a year, fading of up to 20 decibels has been found. This instability was most pronounced from sunset to a few hours after sunrise. The average field strength also tended to be higher during the night. A comparison of stability over various paths from 60 to over 200 kilometers has revealed similar diurnal characteristics. For a shorter path of 30 kilometers, fading was found to be considerably less than for the 60-kilometer path.

The fading magnitude was found to be approximately the same on several 60-kilometer paths differing widely in attenuation and including both optical and nonoptical paths.

Because of the nature of the variations over this 60-kilometer path an increase of 7 decibels in power would have been required to maintain the field equal to or greater than the observed mean value 99 per cent of the time.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

NEGATIVE-ION COMPONENTS IN THE CATHODE-RAY BEAM*

By

C. H. Bachman

AND

C. W. Carnahan

(Hygrade Sylvania Corporation, St. Marys, Pennsylvania)

Summary—*The cathode-ray beams of electrostatically focused cathode-ray tubes were studied by mass spectrographic methods. The negative-ion components focused on the fluorescent screens were made evident by their ability to change the fluorescent qualities of the screen material. Persistent appearance of m/e values of 16, 26, 35, and 37 in the tubes studied were due to oxygen, some organic molecule, as C_2H_2 or CN, and chlorine, respectively. Many other components were found in tubes with various preparation histories. Direct evidence of negatively charged molecules of the cathode material were found which may be of some aid in the theory of sputtering. An explanation of screen discolorations appearing under different conditions of bombardment and life is given in terms of the effects of negative ions, positive ions, and electrons of different energy. The presence on the bombarded screen of a latent image which may be rendered visible by chemical means is discussed.*

ONE of the difficulties encountered in the manufacture of most electrostatically focused cathode-ray tubes for television reception is the appearance on the screen of the so-called "black spot." This blemish appears after periods of operation of the tube ranging from a few minutes to many hours, and consists of an area on the screen which has roughly the shape of an undeflected, focused electron spot and which either has a color unlike that of the screen or is merely blackened. The length of time it takes to become objectionable depends upon the conditions under which the tube was made, the type of fluorescent screen, and the operating voltages. It has been generally understood that the discoloration is caused by negatively charged heavy particles.[1,2] This is easily deduced from the fact that these particles respond to the electrostatic fields of the focusing system but are not noticeably affected by the magnetic scanning fields.

* Decimal classification: R388. Original manuscript received by the Institute, December 4, 1937. Presented before Rochester Fall Meeting, November 9, 1937.

[1] M. von Ardenne, "Negative Jonenstrahlen bei der Formierung von Hochvacuum Elektronenstrahlröhren," *Arch. für Elektrotech.*, vol. 29, page 731, (1935).

[2] L. Levy and D. W. West, "Fluorescent screens for cathode-ray tubes for television and other purposes," *Jour. I.E.E.* (London), vol. 79, pp. 11–25; July, (1936).

In an electrostatic field, the deflection of a charged particle depends upon the ratio of its charge to its kinetic energy at the point of deflection. From this it follows that if an electron beam is focused electrostatically in a cathode-ray tube, all ion components of the beam will go through the same focus, regardless of their charge or mass. In a magnetic field, the deflection depends upon the ratio of charge to momentum at the point of deflection, and ions of different mass and charge do not follow the same trajectories. The resulting deflection is then proportional to the square root of the ratio of charge to mass. In the study of charged heavy particles using the typical mass spectrograph, the particles having some velocity distribution are passed through electrostatic and magnetic fields in such a way that all particles having the same e/m ratios are selected and focused regardless of their velocities, and a spectrum is obtained wherein each component corresponds to particles having a certain e/m value. If all the particles are focused and have the same velocity, only their passage through a magnetic field is necessary in order to give an e/m spectrum. In the electrostatically focused cathode-ray tube, these conditions are met, the only requirement being a magnetic field of sufficient strength to deflect the negatively charged components into a usable spectrum. For instance, the field required to deflect an atom of oxygen 1 centimeter is about 172 times as great as that required for an electron. This paper is a preliminary report of a study of negative-ion components present in the cathode-ray beam.

The tubes tested were arranged with the poles of a U-shaped electromagnet close to the neck of the tube and as far from the gun as possible. With a direct current of about 12 amperes to the coils, the magnet used gave a spectrum of black spots extending radially for 4 or 5 inches.

The identification of the components can be accomplished by any one of several methods. An electron beam can be used for calibration and the magnet current necessary for its deflection can be used to calculate the current necessary to deflect a given ion a certain distance. However, this requires a knowledge of the characteristics of the field for magnet currents having a difference of several hundredfold. The actual field strength and distribution can be measured by well-known methods such as with a flip coil, and the deflection for a given ion calculated from these data. By proper preparation of a tube, known components can be introduced and by their presence and position the spectral region can be calibrated. Also, if a spectrum of several spots is available, it is possible to work out the dispersion curve empirically being aided by the information available concerning the components due to a knowledge of the materials in the tube and the fact that the

lighter singly charged ions will have masses which are whole numbers.

With the exception of the flip coil, all these methods were used, and as the results checked very nicely, it was not necessary to resort to any unusual refinements in any of the procedures. No doubt the success of all these methods was due to the fact that the effective field and thus the dispersion was subsequently found to be practically linear.

The illustrations are photographs of fluorescing screens. In order to see many of the components, it was necessary to operate the tube at such a low screen potential that in some cases, the fluorescence was just barely visible. This made the task of getting good pictures quite difficult. Because of this and of inescapable losses in printing some of the details discussed will not be clearly discerned in the illustrations.

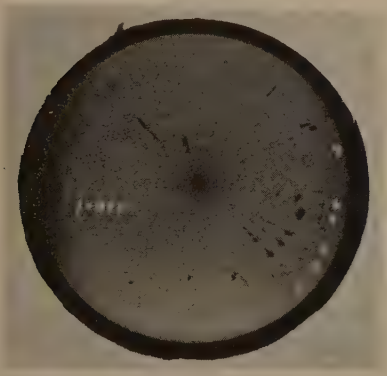


Fig. 1—Cathode broken down after sealoff.

Fig. 2—Cathode aged before sealoff.

Fig. 1 is for a tube which had an oxide-coated cathode. After baking on exhaust, the cathode was brought to a temperature high enough to eliminate most of the gas, but not high enough to break down the cathode.

After sealing off, the cathode was heated sufficiently to start breakdown and give emission. Under these conditions, the spectra obtained were as shown in the figure.

(1) Short exposure for preliminary check
(2) 300 microamperes, 6000 volts, 10 amperes direct current, 10 minutes
(3) 300 microamperes, 6000 volts, 10 amperes direct current, 10 minutes
(4) 300 microamperes, 6000 volts, 10 amperes direct current, 5 minutes
(5) 300 microamperes, 4000 volts, 9 amperes direct current, 10 minutes

Spectrum (5) is of interest in that it shows the effect of lowered second-anode potential and lowered magnet current. The focus was poor in this case and was corrected in spectrum (6).

(6) 400 microamperes, 4500 volts, 9 amperes direct current, 15 minutes

This spectrum was the richest in data of any obtained, 15 components being measured.

The components as taken from (6) are as follows starting at the outside:

m/e	due to
12	C carbon singly charged
16	O oxygen singly charged, O_2 doubly charged, or CH_4 (methane) singly charged
18	H_2O water vapor singly charged or possibly second isotope of O having mass 18 (however, not probable)
19	F fluorine single charged or H_3O (not probable)
26	CN or C_2H_2 (acetylene) singly charged (note ring and center spot)
30	NO (nitric oxide) or HCHO (formaldehyde) singly charged
32	O_2 singly charged
35	Cl singly charged
40	Ca singly charged
56	CaO (calcium oxide) singly charged
74	$Ca(OH_2)$ (calcium hydroxide) singly charged
101	$CaCO_3$ 100 or SrO 103 too small dispersion and too faint to decide
120	SrO_2 (strontium peroxide) singly charged
169	BaO_2 (barium peroxide) singly charged
261	$Ba(NO_3)_2$ (barium nitrate) singly charged

It must be remembered that the cathode in this tube had been only partially broken down. Admitting the possibility of impurities in the carbonate coating and considering the lacquer binder used in suspending the material, all of these ions conceivably could have originated in the cathode region. However, the components due to m/e values 35, 30, 26, and 19, might easily have been introduced in the preparation of the parts. The presence of the heavy components, which are obviously due to the cathode material, is particularly interesting, since this cathode rapidly showed a central pit, due to sputtering. The peculiar ring shown surrounding the spot (m/e = 26) may be evidence, furthering the theory of a gas source other than the cathode.

During the time of obtaining the above runs, the cathode was giving off considerable gas. This is evidenced by the fact that focusing

was more difficult to obtain. As can be seen by other spectra in the figure, focused lines instead of spots were obtained indicating that ionization was taking place along the beam for some distance away from the cathode.

Fig. 2 shows the spectra of a tube identical with the one of Fig. 1 except that the cathode had been broken down and aged for an hour or so before sealing off. To determine the effect, if any, of screen bombardment, the deflecting field was applied before applying the second-anode voltage. Spectra (1) to (9) consist of such runs and show that the components present are

m/e	due to
16	oxygen
26	CN or C_2H_2
35	Cl first isotope
37	Cl second isotope

The centrally located and badly blurred spots are each due to Cl 35 and Cl 37.

At this point, the tube was aged about 20 hours with no second-anode voltage applied, and spectrum (10) shows that the spot $m/e = 26$ has disappeared and $m/e = 16$ has become less intense. At this point, it was noticed that this particular screen was quite sensitive to charged bodies in the vicinity and a grounded metal screen was placed over the face of the tube which seemed to remove this sensitivity. However, with this screen in place the usual spectrum failed to appear. Several more runs made to check this verified the fact.

This can be explained by assuming that with the grounded shield in place the screen potential at the start was too low to allow the negative ions to produce secondaries in greater numbers than the ions. When this happened, the screen went to cathode potential and no more ions were collected. This effect is also observed with very small electron currents with some screens.

Spectra (18) and (19) were made under identical conditions except that before obtaining (19), the screen was scanned for the first time at full rated voltage. One hour of such scanning failed to give rise to any additional components as can be seen in (19).

Fig. 3 shows several points of interest. As can be seen in (1) and (2), several of the spectra are either smeared into lines or not very satisfactorily focused. It was noticed that the cathode temperature was considerably above normal and when another run had been made at lowered heater current, the resolution was much improved as shown in spectrum (4).

The spots appearing on this tube are

m/e	due to
12	carbon
16	oxygen
26	CN or C_2H_2
32	O_2
35	Cl
37	Cl
42.7	measured value, not yet positively identified
80	measured value, not yet positively identified

Fig. 3—Spectra for series of cathode temperatures.

Fig. 4—Effect of gas cleanup by getter action.

On several of these spectra, it can be noticed that the spots due to Cl are surrounded by faint rings. This is pointed out here to be remembered in connection with the similar ring around m/e = 26 in Fig. 1.

The spectra from (7) to (13) consist of a series of runs using heater voltages starting at slightly over normal and extending down to 50 per cent under normal at which insufficient emission was obtained to continue. The grid bias was varied to keep the total current from the cathode the same in each run except the last one (which is very faint) in which only about one half this amount was available. It will be noted that the outermost spot due to carbon, and the spot m/e = 26 gradually become weaker. This may be due to a gradual decrease in the amount of the gas available or to differences in the efficiencies of ionization as the conditions at the cathode change with temperature. Spectrum (12) shows that m/e = 42.7 has changed decidedly and has become smeared out into a large area. No explanation of this is offered at the present time but it is hoped that further study of cathode conditions will bring a better understanding of this phenomenon.

Several of the spectra in this figure exhibit the line focusing earlier mentioned and attributed to ionization along the beam. Spectrum (3) shows this particularly well and when the difference in voltage between the two ends of one of these elongated spots was calculated from the measurement of its length, it was found to be equal to the first-anode focusing voltage. In other words ionization occurred in the beam all the way from the cathode to the first anode.

The long row of spots in spectrum (6), shows a series of runs at different magnet currents, used in obtaining the direct-current field distribution. The exposures were timed so that only the strongest spot appeared and was used for measuring purposes.

Fig. 4 shows the efficiency of clean up of a "Batalum" type getter. The tube was suspected of being a leaker when it did not pull down on the pumps. After sealing it off, however, it was found that the fault had not been in the tube. From the conditions on pumping, it is believed that the pressure in the tube was about 10^{-3} millimeters of Hg. The cathode was heated sufficiently to give emission and an attempt made to secure a black-spot spectrum. The radial streak terminating at $m/e = 16$ is the result. It shows excessive ionization between cathode and first anode and probably clustering with a continuous spectrum resulting. An unflashed "Batalum" getter was then burned out and the several sharp spectra beside the streak were immediately obtained.

In other tubes where desired gases were introduced by heating side tubes, excess pressures were repeatedly reduced to usable values by burning a "Batalum" filament for a few minutes at bright red.

Fig. 5 is for a tube which contained a hairpin tungsten filament instead of the usual oxide-coated cathode. It should be noted that there are several types of rings, some diffuse and some sharply defined at both inner and outer edges. Also some of them have a center spot and a spot in the outer edge of the ring. A tentative explanation of these phenomena has been developed and will be offered after the completion of substantiating experiments now in progress.

The next three figures, 6, 7, and 8, are photographs of tubes which had been on life test and which with lowered screen voltage show a pattern similar to the scanning pattern with which they had been associated. Here again the effect is similar to the black spot in that the screen material seems to have changed its character over the affected area but bombardment with higher voltage electrons makes it unnoticeable.

Fig. 6 shows a pattern consisting of the large scanning pattern, stronger diagonal lines running to the center, a still stronger round area, and contained in this a still stronger but small scanning pattern.

Fig. 7 shows this same effect on another tube with a little better detail in the center.

Fig. 8 is this same tube with a different voltage adjustment making the detail of the larger pattern more clear.

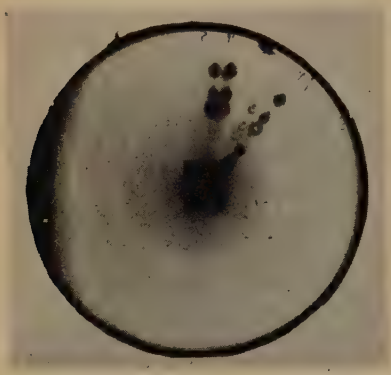


Fig. 5—Tungsten filament type.

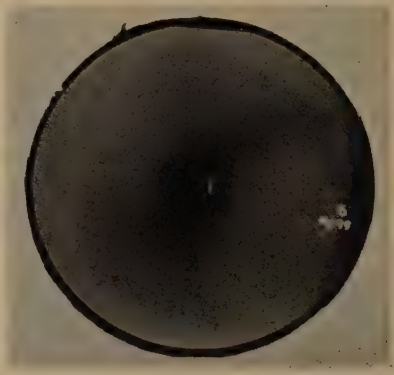


Fig. 6—Life-tested tube showing positive-ion blemishes.

At the present time at least, the following explanation seems to account for this pattern.

In the first place, evidence shows that when a tube develops a life-test pattern such as this, it has become somewhat gassy even though the tube characteristics still permit satisfactory use.

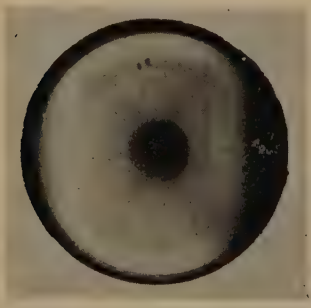


Fig. 7—Positive-ion blemish with more detail at center.

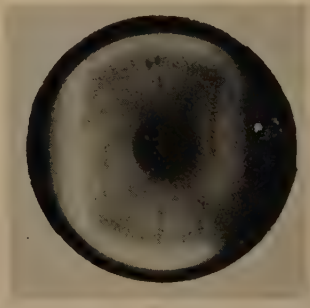


Fig. 8—Same as Fig. 7 adjusted for more detail away from center.

The large pattern probably is due to changed fluorescent qualities resulting from a change in the crystal structure of the screen material after sufficient electron bombardment.

The diagonal lines are due to positive-ion bombardment under the following conditions. When the electron beam is deflected so as to scan

a rectangular or square area on the screen, the volume traversed by the beam through the tube is roughly in the shape of a pyramid. The currents applied to the deflecting coils consisted of sine waves in each case. Under these conditions, the component of the electron scanning velocity in the horizontal is minimum at maximum deflection and similarly with the vertical deflection. Thus at the corners we have the maxima of both deflections with the resultant maximum of time spent there by the electron beam. Since the absolute velocities of all the electrons is the same, if we find the beam spending more time in one region than any other, we might conclude that more ionization by electron collision will take place in this region than others. Now any positive ions formed in the large bulb of the tube will be drawn directly to the screen because this floats at some potential below the second-anode voltage on the conducting walls and is the most negative area in the bulb interior. Therefore, the corners of the pyramid starting at the screen and extending back to the undeflected beam, if projected upon the screen, give lines denoting maximum reception of positive ions formed by the electrons in the scanning pyramid. Over a long period of time, the effect of these ions at the screen is similar to that of the negative ions causing the black spot. The efficiency of fluorescence is altered and a color change may result.

The miniature scanning pattern is attributed to positive ions formed in the neck of the tube between the second-anode—first-anode lens and the deflection coils. The ion beam is of sufficiently low voltage to enable the fields of the deflection coils to give it a noticeable deflection.

The small circular pattern is probably due to positive ions formed in the neck of the tube but on the screen side of the deflection coils. These are pulled directly to the screen and give an image of the neck cross section.

The above explanations are based largely upon the idea of positive ions being attracted by a screen floating at a lower potential than the surrounding walls. This effect can be shown quite easily with the average tube in another way. With the scanning circuits off, an unfocused beam is allowed to fall upon the screen. With other values normal, the second-anode potential is lowered to a value at which the beam suddenly breaks away. If the voltage is now raised slowly, the beam will appear but will contain a dark center patch of irregular size and shape and the fluorescing rim will appear to be shooting long fluorescing streamers across the screen in many directions. This black patch by itself usually shrinks in size until it suddenly closes up, the streamers disappear and the beam is normal. This process can be

speeded up by raising the second-anode voltage or varying the beam current. The black patch is evidently due to that region being so low in secondary emission or space ion current that it cannot maintain itself sufficiently positive to receive the current from the beam. This phenomenon also has been noticed by Koller and Johnson[3] in connection with a study of the Malter effect. One thing not observed by them, probably due to their use of willemite, is the discoloration of the fluorescing screen under such conditions. If a small black patch, which we have shown is more negative than the surrounding screen, is allowed to remain on the screen for a short time, the positive ions attracted to this spot have sufficient energy to form an image of the area, by changing its fluorescing qualities. For instance, using a sulfide screen of blue color, the affected spot fluoresces green. The patterns due to life test discussed above give definite indication of change of color. As a matter of fact the so-called black spots formed on the same screen by negative ions have a dark purple color. It would seem that the difference in energy between the high-voltage negative ions and these positive ions might account for the former giving rise to black or purple spots while the ions of lesser energy give a different color.

In connection with the effect of bombardment of the screen, one rather interesting experiment might be described. A year or two ago, one of the authors working with a demountable system dropped into a tray of developer which happened to be convenient a plate with a screen having a fairly strong black spot. After about an hour in the developer during which time nothing developed out, the plate was slipped into the hypo tray. After some little time in the hypo, it was found that there had developed a physical image of the black spot which had appeared under fluorescence. On the basis of recent work done along these lines, but which is not yet complete, it appears definite that a latent image is formed by bombardment and the results described above were quite reasonable.

Returning to the appearance of negative ions in the electron beam of high vacuum tubes, it should be pointed out that, although it has been known for some time[4,5,6] that an oxide-coated cathode is a source of oxygen, there has been a definite lack of agreement as to its carrying a charge. It seems conclusive that the oxygen is, in some cases at least,

[3] L. P. Koller and R. P. Johnson, "Visual observations on the Malter effect," *Phys. Rev.*, vol. 52, pp. 519–523; September 1, (1937).

[4] F. Detels, "Ueber den Formierungsprozess in Oxydkathodenrohren," *Jahr. der Draht. Tel. und Tel.*, vol. 30, pp. 10–52, (1927).

[5] J. D. Becker, "Phenomena in oxide-coated filaments," *Phys. Rev.*, vol. 34, pp. 1323–1351; November 15, (1929).

[6] H. A. Barton, "Negative-ion emission from oxide-coated filaments," *Phys. Rev.*, vol. 26, pp. 360–363; September, (1925).

negatively charged and the detection of these ions is probably determined largely by the arrangement of the apparatus and the fields used. In a recent article[1] Levy and West conclude that the negative-ion black streak on their tubes was due to atomic and molecular hydrogen ions. Although this region of the ion spectra was searched carefully, no trace of an ion of m/e value less than 12 was found.

The exact mechanism of cathode sputtering is not clearly understood at the present time. The spectra in Fig. 1 show definitely components of the cathode material leaving the cathode region as charged molecules. The cathode of this tube was sputtered away at a rapid rate. This experiment does not rule against the formation of charged molecular aggregates during the sputtering, since their mass would have been too great for noticeable deflection, but it does indicate the presence in the sputtered components of charged single molecules.

It should be noted that the results shown from this preliminary study indicate that much can be done along these lines in the study of thermionic emission, negative-ion formation, and the phenomenon of sputtering. The positive ions undoubtedly present in the tubes could not be studied with the same experimental setup but with the necessary modifications for their measurement, a fairly complete picture of the processes at the cathode should be revealed. Further work in these fields as well as the effect of screen bombardment is under way.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

THE FINE STRUCTURE OF TELEVISION IMAGES*

By
Harold A. Wheeler and Arthur V. Loughren
(Hazeltine Service Corporation, New York, N. Y.)

Summary—*There are described the necessary or preferred relations in the fine structure of television scanning, so chosen as to yield a system which is both theoretically ideal and practically optimum. The relations involve the width and spacing of the lines and the intensity distribution of circular scanning spots in both transmitter and receiver, which are correlated with the filter characteristics and the required width of the frequency band. A "flat" field of scanning, as distinguished from a "half-tone" field, is required at both ends, especially in the transmitter to suppress spurious large patterns otherwise generated in the scanning process. For equal horizontal and vertical "width of confusion," the nominal cutoff frequency of the system is found to be*

$$f_c = f_p N_s^2 R / 2\sqrt{2}$$

in which

f_p = *frame frequency*
N_s = *number of scanning lines, and*
R = *aspect ratio.*

The over-all electrical characteristics require distortionless reproduction up to f_c and permit a gradual cutoff between f_c and $2f_c$. The "quadratic-sum rule" is derived for a special case and is offered as a generally approximate expression of the image width of confusion w in terms of the width of confusion of the respective individual components (object, optical, scanning, electrical):

$$w = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}.$$

Based on the proposed television system of 30 frames per second, the full utilization of 441 lines requires a separation of 5.5 megacycles between picture and sound carriers. A separation of 4.25 megacycles is suggested as a practical compromise within the proposed 6-megacycle channel.

I. Introduction

THE television systems which appear to be realizable practically at the present time are systems which, at any given instant of time, transmit a single picture element. They depend on a scanning process for image analysis at the transmitting camera and for image synthesis at the receiver. The scanning process is an artifice which permits the number of communication channels needed for satisfactory reproduction of an image to be reduced from perhaps one hundred thousand down to a single channel.

* Decimal classification: R583. Original manuscript received by the Institute, January 10, 1938. Presented before Cincinnati Section, November 16, 1937; Emporium Section, November 17, 1937; Indianapolis Section (in cooperation with the American Association for the Advancement of Science), December 28, 1937; Connecticut Valley Section, January 6, 1938; Chicago Radio Engineers Club, February 17, 1938, by J. K. Johnson; and New York meeting, April 6, 1938.

Since scanning introduces elements of distortion which are not present in the camera image, these affect the reproduced image. It is essential that the influence of the scanning process on the image-quality and on the communication-channel requirements be understood, in order to make available a sound basis for the establishment of a practical television broadcast system.

Perhaps the full importance of the fine structure of the images has not been generally appreciated because the observer should be far enough from the reproduced image so the fine structure is not obvious. The outstanding exception is the spurious large pattern such as described by Mertz and Gray, which remains even when the scanning lines are not resolved, in spite of the fact that its cause lies in the fine structure of the transmitter scanning, specifically in the space between adjacent scanning lines. There has been especially the need for a more simple theoretical treatment of the conditions for minimizing such major distortion, and the resultant vertical resolution as determined by the optimum transverse characteristics of the scanning lines of the transmitter and receiver.

Most of the literature on this subject pertains to the effect of the size and shape of an aperture having a discrete boundary, as utilized in mechanical scanning, the light intensity being uniform within the boundary. When the scanning spot is obtained by an electron beam, a more practical distribution of intensity is a circular distribution having maximum intensity in the center, gradually decreasing with increasing radius. The factors leading to an optimum circular distribution have not been treated.

An important factor that must be standardized for regulatory purposes is the width of frequency band required. The analogous problem in engineering work is what requirements are placed on the design of electrical transducers for television signals. There has been the need for a more comprehensive theoretical determination of the band width required, with special attention to the minimum permissible frequency separation of adjacent picture and sound carriers.

The analogy between the confusion in the fine structure of a television image, and the familiar "circle of confusion" in optics and photography, is so close that similar terminology should be used. The scanning process makes this confusion "two-dimensional," that is, the horizontal and vertical confusion depends on different factors. Therefore, instead of using the circle of confusion as the criterion of resolution, it is preferable to use horizontal and vertical "width of confusion." Just as the circle of confusion (c.o.c.) is based on the image of a point object, such as a star, the width of confusion (w.o.c.) is based on the

image of a very narrow line, such as a straight incandescent filament or the slit of a spectroscope. The horizontal width is based on the image of a vertical line. The vertical width is based on the average image of a nearly horizontal line which is cut by the scanning lines at a very small angle. If the horizontal and vertical widths are equal, the equivalent circle of confusion is the diameter of a circular spot which would generate lines of this width. This diameter is about 1.2 times each width, so this conversion factor may be used in comparing resolution expressed in these different terms.[1]

The analogy between a circular aperture in television and a circular iris in a pinhole camera is exact, as is also the analogy between the granular structure of a camera-tube mosaic or a fluorescent screen and that of a photographic deposit. In optics, there is nothing like the low-pass electrical transducers of television, but there is the ultimate limitation on the resolution of a microscope, which is imposed by its diameter as compared with the wavelength of light. Therefore it is interesting to compare the band width of a filter with the lens diameter in optics, a greater value of either serving to reduce the width of the resulting diffraction pattern in the image. Of course, there are also the parts of the television system which involve optics directly, namely, the optical system of the camera (and the projector, if one is used in the receiver) and that of the human eye.

In the previous theoretical treatment of television images and the limit of resolution, it has been customary to use as a criterion the sharpness of reproduction of the edge of a flat field or a stripe of appreciable width. This object is not as sensitive to distortion as a narrow bright line, and the line is much simpler to analyze, so it is chosen as the basis of the present treatment. It shows forms of distortion not otherwise appreciated, such as "beads" which form spurious patterns of major size superimposed on the desired image. Also a single line is mathematically a more critical test than a large number of lines evenly spaced, even though not so well adapted for visual observations.

Previous treatments have therefore dealt with the requirements for reproduction of finite objects with appreciable distortion. This study instead deals with the "width of confusion" contributed by the system in the reproduction of a narrow line having no appreciable width of its own, so the resulting values are independent of the nature of the object.

[1] The ratio of the c.o.c. to the w.o.c., for a circular distribution in the scanning spot, has a value depending on the radial variation of intensity. The two extreme cases of those likely to occur in practice are the discrete circular aperture and the probability curve, for which this ratio is, respectively, $4/\pi = 1.27$ and $\sqrt{4/\pi} = 1.13$. The mean of these values is 1.20, the value given above. The ratio is 1.17 for the ideal circular spot to be derived (Appendix B).

One of the main requirements of the fine structure is to minimize spurious images much larger than the area of confusion. This is accomplished by relating the width and pitch of scanning lines to secure a "flat field" of scanning in the transmitter, regardless of the receiver. It is also desirable to secure a flat field of illumination in the receiver, so the individual lines cannot be discerned.

Then a lateral distribution of intensity across the scanning lines is chosen which meets these conditions and which can be obtained with a moving spot of reasonable circular distribution. All of these factors together determine uniquely an ideal distribution for the lines and spot, which in turn determines the vertical width of confusion. This is the average width of the final image of a narrow line nearly horizontal.

As the ideal circular spot scans a vertical line, the horizontal distribution across the vertical line image is nearly the same as the vertical distribution across a scanning line, so the resulting horizontal width of confusion of the scanning system is nearly equal to the vertical. This leaves little tolerance of filter distortion in the electrical circuits, so there remains only to determine what is the least frequency band that must be passed and what is the sharpest cutoff characteristic that is permissible. Phase distortion is assumed to be absent, because it is not an essential limitation and can be compensated to any degree of approximation.

The Fourier integral analysis of the transient wave form produced by scanning spots shows that the nominal cutoff frequency of any system is that at which the width of a half cycle equals the nominal value of the horizontal width of confusion (as here to be defined). It also shows that a filter with a sharp cutoff at that frequency makes the real width of confusion much greater because it widens the main pulse of the transient and also adds spurious positive and negative pulses like a diffraction pattern. Therefore it is preferable to have a gradual cutoff between the nominal cutoff frequency and twice that frequency, which avoids too much widening and makes the spurious pulses negligible.

This study yields a complete theoretical derivation of a formula properly relating the factors of scanning with the frequency characteristics of the low-pass filters. In order to obtain the desired gradual cutoff, the television signal should have sufficient frequency separation between the picture carrier and the nearest other signal (on at least one side). This separation should be twice the nominal cutoff frequency of the system. For example, the full utilization of the present 441-line system requires 5.5 megacycles separation between the picture carrier and the nearest sound carrier on one side. Not so much separation is needed on the other side, since only one side band has to contain all the information in the signal.

II. The Vertical Resolution

The vertical resolution is reciprocally related to the vertical width of confusion, and the latter is a more convenient quantity to use in this study. These are determined by the transverse characteristics of the horizontal scanning lines. The width of the frequency band of the electrical system places no restriction on the vertical resolution because this involves only components of relatively low frequencies, of the order of the line frequency, whereas the horizontal resolution involves components of much higher frequencies.

The transverse characteristics of the scanning lines are the pitch of the lines and the transverse distribution of intensity. The former has a simple numerical value which depends only on the scanning system, whereas the latter requires a functional representation and depends on

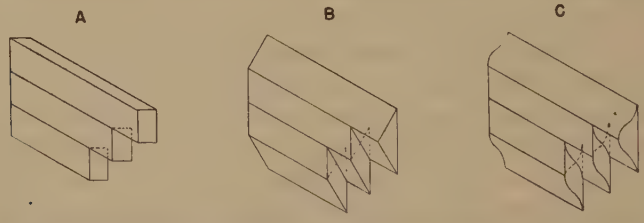


Fig. 1—Transverse characteristics of scanning lines to produce a flat field.

several factors. In the transmitter, it depends on such factors as the optical system forming the image on the camera screen, the grain size of the screen surface, and the diameter and shape of the aperture or the intensity distribution of the scanning spot. In the receiver, it depends on the corresponding factors in reverse order, of which the optical system includes the eye of the observer. The effective transverse distribution is really the combined effect of such factors, but for present purposes it is convenient to consider only the scanning spot, since this is the only factor essentially involved.

It is desirable in a television receiver that the picture screen show a flat field when reproducing a flat field, that is, that the line structure be indistinguishable. This places no restrictions on the scanning spot at the transmitter but requires that the scanning lines at the receiver appear to have sufficient width and the proper amount of overlapping. This requirement fades in importance as the observer moves farther from the screen, but so do most of the factors influencing the fine structure, so this is no reason for neglecting this factor.

By way of example, Fig. 1 shows, in three-dimensional figures, blocks of several scanning lines. The relative brightness is shown as the

"thickness" dimension. In Fig. 1A, the lines have uniform intensity, as obtained by a rectangular aperture, and are contiguous without overlapping. A uniform field results since the line width equals the line pitch. (This result would be obtained also if the width were any integral multiple of the pitch, because sets of nonadjacent lines would be contiguous even though adjacent lines would overlap.) In Fig. 1B, the lines have triangular distribution, as obtained by a diamond-shaped aperture whose lateral diagonal is twice the pitch. The overlapping is just sufficient to secure a flat field. In Fig. 1C, the lines have cosine-squared distribution which amounts to rounding out the triangular distribution while retaining the flatness of field.[2]

Two factors influence the choice of an ideal lateral distribution. One is the criticalness of overlap and the other is the shape of spot distribution required. The former is most severe for the rectangular shape, least severe on the sides for the triangular shape, and least severe on the edges for the cosine-squared shape. A choice between the second and third cases is influenced by the spot requirements.

The most practical scanning spot is one having circular distribution of intensity, with a rounded peak at the center, sloping sides, and flat edges, because this shape is naturally secured by means of an electron beam focused by an electron-lens system.[3] Any other shape has less tolerance of errors of focusing. This shape is similar to a circular aperture in "soft focus," and may be obtained by just such an expedient in the cathode-ray system.

The ideal lateral distribution of the line must therefore be chosen with a rounded peak of more curvature than a semicircle, because semicircular distribution is obtained from a circular aperture in sharp focus. The edges must be flat to fit the peaks of adjacent lines. The simplest continuous function which meets these requirements and gives a flat field is the cosine-squared

$$v = \cos^2 \frac{\pi}{2} u = \frac{1 + \cos \pi u}{2} \tag{1}$$

in which v is the relative intensity across the line and u is the distance from the center relative to an assumed value of unity for the line pitch. The area under the distribution curve for one line is unity, represent-

[2] Mertz and Gray (12) show these three cases in their Figs. 23, 25, and 26, respectively. Their Fig. 27 shows the third case with 3/2 as many lines, which also gives a flat field and is less critical as to overlapping, but which ultimately requires 3/2 as wide a frequency band to transmit the same amount of information. (All figures in parentheses in footnotes refer to the bibliography.)

[3] Zworykin (11), in his Fig. 7, shows this shape of radial distribution of a spot. Jesty and Winch (19), in their Figs. 13 to 15, show similar lateral distribution of lines (note that the ordinates of Fig. 14 are on a logarithmic scale).

ing its contribution to the flat field. The entire width of each line is twice the pitch, to fill in between the central ridges of adjacent lines. It cannot be any less, but can be a greater integral multiple of the pitch. Therefore the number of scanning lines may be increased in the ratio of any half integer while maintaining the flat field. This expedient, however, increases in the same ratio the width of frequency band required to transmit the same amount of information.

There is a more important defect caused by improper transverse distribution of the scanning lines in the transmitter. This defect is the appearance of "beads" on the image of a line nearly horizontal. The beads become longer, and therefore discernible when viewed at greater distances, as the line image becomes more nearly horizontal. On the image of many fine lines closely spaced, even though the lines are not distinguishable in the image, the beads form a spurious major pattern superimposed on the desired image.[4]

On closer study, this spurious pattern is found to be caused most severely in scanning a "half-tone flat field," that is, a field which appears to be flat but is really composed of many fine lines close together. It is caused by the failure of the transmitter scanning system to respond equally to all parts of the fine lines. The obvious solution lies in the direction of blurring the response to the fine lines in order to secure the effect of an object field which is really flat instead of half tone, and this may be accomplished by optical defocusing or by widening the scanning lines in the transmitter, so the half-tone lines cannot be resolved. The problem therefore assumes the same aspect as the flat-field requirement in the receiver, and its solution lies in the requirement of a flat field of scanning in the transmitter.

Fig. 2 shows in elementary form how beads may be produced in the transmitter scanning, and how the receiver cannot cause them or compensate for them. For simplicity, there is assumed in both transmitter and receiver a rectangular aperture very narrow in the horizontal direction but of various lengths across the scanning lines. Two adjacent scanning lines in the transmitter and the corresponding two in the receiver are shown superimposed, and their lateral distributions (widths) are shown in cross section at the right-hand side. A very narrow bright line L is assumed as the object and is shown crossing the scanning lines at a small angle from the horizontal. The duration of the signal transmitted on each scanning line depends on the line width in the transmitter, between the dashed lines in the drawing. The receiver reproduces

[4] Mertz and Gray (12) show this phenomenon in their Fig. 18 and give an elegant mathematical theory of its production, but fail to describe in simple terms its cause and the requirements for its elimination.

the signal as a rectangular area of the same duration and of the same width as the receiver lines, between dotted lines in the drawing. Fig. 2A shows the simplest case of contiguous nonoverlapping lines in both transmitter and receiver, B and C show respectively narrower and wider lines in the transmitter only, and D shows narrower lines in the receiver only.

In the interpretation of Fig. 2, the vertical "steps" on the edges of the images should be overlooked, because they are unavoidable with the assumed aperture. The pattern may be viewed from a distance

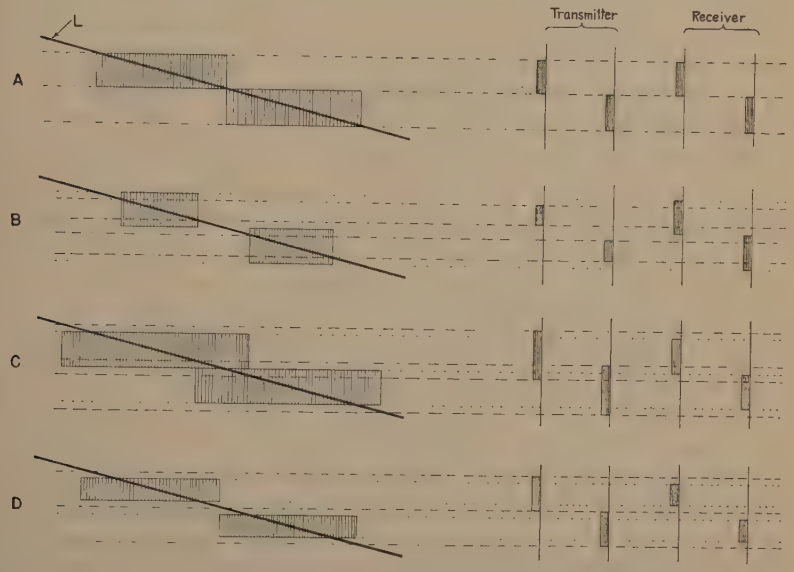


Fig. 2—The beads caused by failure to have a flat field of scanning in the transmitter.

sufficient to make these steps disappear, but not sufficient to make the horizontal irregularities disappear. Then the image of Fig. 2A appears to be a sloping line of uniform intensity along its length, B appears as a dashed line, C appears as a string with beads spaced along its length, and D appears also as a uniform line because the adjacent image areas fail to meet only by a distance comparable with the line pitch. These observations become more definite as the slope of the object line is decreased, whereas this slope is exaggerated in the drawing. The B and C cases, in practice, both appear as beads, but in B the visibility of the string depends mainly on the transverse distribution of the transmitter lines. The essential difference between A and D on one hand, and B and C on the other hand, is that the former but not the latter have a flat

field of scanning at the transmitter (as in Fig. 1A), so the latter show the appearance of long beads caused by nonuniform intensity of light along the image of the nearly horizontal line. As the slope of the line increases, the beads become shorter until their size is comparable with the line pitch and with the steps characteristic of the assumed apertures.

The conclusion from the study of Fig. 2 is that a flat field of scanning in the transmitter is of even greater importance than in the receiver, because its failure causes the large spurious patterns which are

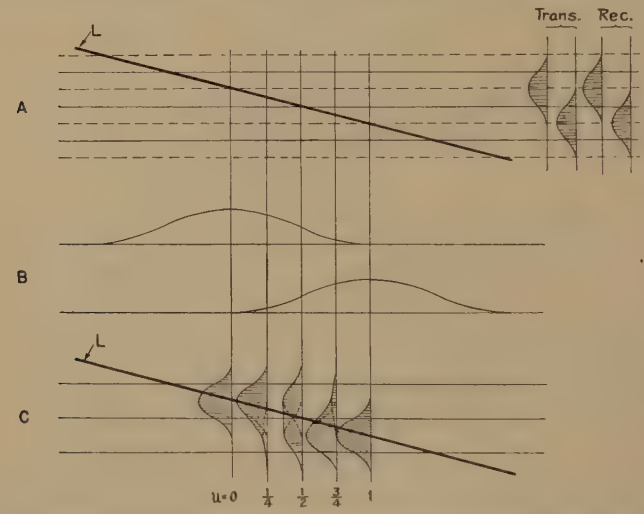


Fig. 3—Derivation of the vertical width of a nearly horizontal line image.

easily seen, even when viewed at distances so great that the fine structure is unobvious. The choice of lateral distribution of scanning lines in the transmitter depends on factors corresponding to those involved in the receiver, so the same lateral distribution is chosen as ideal. In a camera tube with cathode-ray scanning, the spot distribution depends on the diameter of the focused electron beam, just as it does in a picture tube, because this distribution determines the relative response over the surface of the screen, even though it is not visible as a spot.

The vertical resolution of the system depends on the transverse characteristics of the scanning lines of both transmitter and receiver. It is inversely related to the vertical width of confusion, which is to be derived for the case of the ideal lateral distribution.

Fig. 3 shows the analysis of the image of a narrow line nearly horizontal, as in Fig. 2. A flat field of scanning is obtained at both ends of the system by the cosine-squared distribution of Fig. 1C, which has

been chosen as the ideal. Fig. 3A shows the lateral distribution in cross section at the right-hand side, and shows the object line L crossing the scanning lines at such a small angle that limitations of horizontal resolution can be neglected. Fig. 3B shows the signal carried by each scanning line in the transmitter, which is the variation of intensity at the intersection of the object line with the scanning line. This variation is merely a projection of the lateral distribution curve, so it has the same cosine-squared shape. This is the relative amount each scanning line in the receiver contributes along the final image. Fig. 3C shows the resulting lateral distribution in five cross sections along the line image, denoted by successive values of u, the relative distance of the object line from the center of one scanning line to the center of the next.

The transverse characteristics of the line image vary along the line. The width of the image varies between one and two scanning lines, having the former value where the object line crosses the center of a scanning line, and the latter value where it is midway between two adjacent scanning lines. The vertical width of confusion is the width of this line image, taken as an average along the line.

The definition of the width of a scanning line or of a line image is somewhat arbitrary because it is assigning a number to represent a curve such as the cross sections of Fig. 3C, one of which is shown in Fig. 4 for further analysis. The usual expression for the width of a curve is the ratio of the area to the peak value, that is, the width of a rectangle having the same area and height. In this case, however, the peak value has less significance because it does not generally occur where it is most desired, namely, where the line image should be. Therefore, the width is here defined as the width of a rectangle having the same area as the curve of relative intensity, but whose height is the height of the curve over the point where a perfect line image would pass under the curve.[5] This rectangle, of width w, is shown dotted in Fig. 4, with a dotted dividing line at the point u where the line image should be.

Appendix A is the mathematical derivation of the average width of the line image, which is the vertical width of confusion. Its value is found to be $\sqrt{2}$ times the pitch of scanning lines.[6] Therefore the

[5] This definition, in this case, properly penalizes the transverse characteristic where its peak is not at the most favorable position. Even this method, however, would fail to penalize the distance off-center of the transverse characteristic of Fig. 2A, or of others obtained by a rectangular aperture in the receiver. The reason is that it fails to penalize inequality of the partial areas on either side of the dividing line at u.

[6] This is a special application of the quadratic-sum rule to be described, in which it happens to give the exact value. In other words, the vertical width of confusion of the system is the quadratic sum of the line widths in transmitter and receiver.

vertical resolution is the number of scanning lines, divided[7] by $\sqrt{2}$.

So far, little attention has been paid to the circular distribution of intensity in the scanning spots, as determined by the various factors involved, such as the diameter of the electron beams. There remains to determine the radial variation of intensity of a scanning spot which, when moving uniformly across the screen, generates a line having the

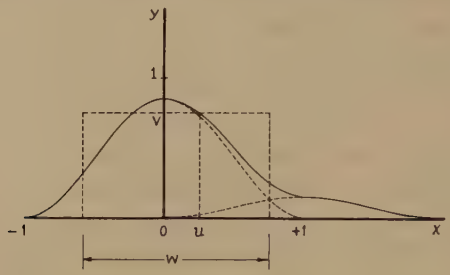


Fig. 4—Definition of the width of a line image.

ideal lateral distribution. This is the subject of Appendix B, in which there is derived the radial distribution formula

$$z'/z_0' = 0.675(1 - r^2)^{3/2} + 0.325(1 - r^2)^{7/2} \qquad (2)$$

in which r is the radius (relative to unity pitch of lines) and z'/z_0' is the intensity relative to its maximum value at the center. This formula is not exact but is based on a very close approximation. The derived

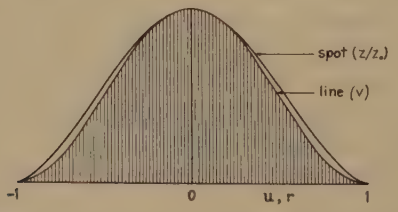


Fig. 5—Distribution curves of the ideal scanning line and the circular spot required to produce it.

radial characteristic of the ideal spot is shown in Fig. 5 together with the ideal transverse characteristic of the resulting scanning line. The effective diameter of the spot is 1.17 times the line width, which is required because the side portions of the spot are effective for a shorter distance and time than the portion along the center of the scanning

[7] This gives the theoretical value $1/\sqrt{2}=0.71$ for the correction factor which has been experimentally estimated to lie in the range of 0.64 to 0.75 (bibliography (13) and (20)).

line. The derived distribution is of a form that can be secured practically to a very close approximation.[8]

III. Horizontal Resolution

The horizontal resolution is reciprocally related with the horizontal width of confusion, and the latter is a more convenient quantity to use in this study. These are dependent mainly on factors which do not directly affect the vertical resolution, such as the horizontal width of the scanning spots and the electrical low-pass filter characteristics. The use of a circular scanning spot, however, implicitly relates the vertical and horizontal widths of confusion of the scanning system, making them nearly equal. This approximate equality is subject to verification as the resultant effect of the two scanning systems of transmitter and receiver. The resultant vertical width of confusion has been derived, so there remains to derive the horizontal resultant of the two systems, employing circular scanning spots of ideal distribution.

The horizontal width of confusion of the two scanning spots of the transmitter and receiver is the horizontal width of the image of a narrow vertical line object. In Appendix C, using the ideal circular scanning spots, this width is found to be 4/3 the pitch of scanning lines. This is only 6 per cent less than $\sqrt{2}$, the vertical width of confusion resulting from the use of the same scanning spots.

It is desired to secure equal horizontal and vertical width of confusion, but only the horizontal width is effectively increased by the electrical filters. Therefore these filters should have such characteristics that they add only about 6 per cent to the horizontal width of confusion caused by the two scanning spots. Also they should not cause appreciable spurious pulses in addition to the desired central pulse.

The horizontal width of confusion is the nominal width of an image pulse which is viewed in space but which is caused by a signal pulse that is transient in time. The relation between transient signals and the frequency characteristics of electrical filters is analyzed by the use of the Fourier integral. The properties of the Fourier integral, and the results obtained by its application to the present problem, are treated mathematically in Appendix D. They are summarized here with reference to Fig. 6, in order to form a theoretical basis for setting up the requirements of the filter characteristics.

In this discussion, the use of the Fourier integral is greatly simplified at the outset by ruling out electrical phase distortion. It plays no

[8] Zworykin (11) in his Fig. 7 shows a curve of similar shape representing the observed radial distribution of a scanning spot. Also Jesty and Winch (19), in their Figs. 13 to 15, show similar shapes of observed lateral distribution of scanning lines.

essential part in the analysis, since it can be compensated to any degree of approximation by phase-correcting networks. Phase distortion is the cause of a symmetrical input transient being reproduced as an unsymmetrical transient in the output. The elimination of phase distor-

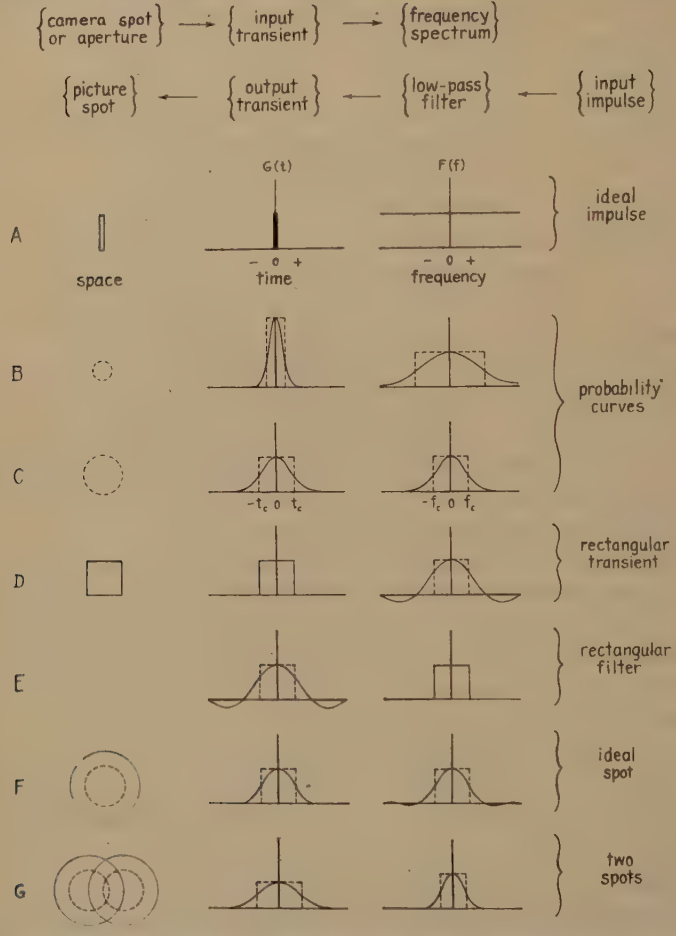


Fig. 6—Relations between a scanning spot, the corresponding transient signal, and its frequency spectrum.

tion from consideration simplifies the analysis by restricting it to the symmetrical reproduction of input transients. Only symmetrical input transients are considered, obtained by horizontally scanning a narrow vertical line object with scanning spots having horizontal symmetry. Therefore all the transient signals to be considered, both input and output, are pulses which are symmetrical on the time axis. They are repre-

sented by curves showing the time variation of intensity, and are interpreted as the horizontal space variation of intensity in the resulting image.

In comparing a transient wave form with its frequency spectrum, it is helpful to regard the latter as symmetrical on the frequency axis. Any frequency characteristic is inherently symmetrical about zero frequency because negative and positive frequencies of equal magnitude are not physically distinguishable. In a frequency spectrum, the positive and negative frequency bands may be regarded as the side bands produced by modulating a zero-frequency component analogous to a carrier wave.[9] If this spectrum is employed to modulate a high-frequency carrier, these side bands then become separable, because the symmetrical frequency spectrum is shifted bodily along the frequency axis.

By limiting the analysis to symmetrical transients, and by representing the frequency characteristics also as symmetrical, there are obtained pairs of curves such that either one of a pair represents a transient wave form whose frequency spectrum is given by the other one of the pair. The Fourier integral in this case involves not the sine but only the cosine function, which is symmetrical about zero frequency.

The nomenclature of the Fourier pairs corresponds to that of Campbell and Foster. The frequency function is denoted $F(f)$ and the time function is denoted $G(t)$. If either curve of a pair represents one function, the other curve represents the other function.

The extreme case in the application of the Fourier integral is the instantaneous impulse and the corresponding uniform frequency spectrum, shown in Fig. 6A. This is the limiting case of the probability curves of Figs. 6B and 6C. The probability curve is the unique case in which the two curves of a pair have the same shape. The two are shown identical in Fig. 6C, and in Fig. 6B are shown modified in the direction of the limiting case of Fig. 6A. The ideal impulse of Fig. 6A has finite area, obtained by very large amplitude and very small duration, so its shape is immaterial, but it is logically regarded as an extremely narrow probability curve. Conversely, its frequency spectrum is logically regarded as an extremely wide probability curve. The converse meaning of this pair of curves is that a signal of uniform intensity (such as a direct current) has zero width of frequency spectrum.

[9] If the components of a Fourier series, representing a periodic function, are disposed symmetrically about zero frequency, the same formula can be used to compute all terms, whereas otherwise the formula for the zero-frequency term has a coefficient half as great as that of the formula for the other terms. The unified formula for the symmetrical components is the basis on which the Fourier integral is derived by passing to the limit of zero fundamental frequency.

The nominal width of one of these curves is the width of a rectangle having the same height at the center and having the same area, the area under the curve being the net value of positive and negative areas. The width of a transient is denoted $2t_c$. That of a frequency spectrum is denoted $2f_c$, in which f_c is the nominal cutoff frequency. The Fourier integral yields the exact relation that the product of these widths is unity. In other words, the transient width is half the period of one cycle at the cutoff frequency. These quantities are precisely defined, but they are not physically absolute because the curves extend beyond their nominal widths.

It is desirable to have both curves cut off as sharply as possible beyond their nominal widths. The edges of the transient cause confusion with the adjacent regions of the picture. Those of the frequency spectrum include components which would be rejected by a low-pass filter with sharp cutoff in the vicinity of f_c. The probability curve falls rapidly outside the nominal width, but never reaches zero absolutely. Any sharper cutoff of one curve causes more gradual cutoff of the other, so it is apparent that the probability curve is the theoretically ideal form, in which both curves of a pair have equally sharp cutoff.

If either curve of a pair goes to zero beyond a certain width, the other acquires spurious pulses of negative or positive area. For a given nominal width, these cause the central pulse to have greater or less positive area. Therefore the effective width, which depends rather on the absolute sum of positive and negative areas, may be increased by a complete cutoff, to a value greater than the nominal width.

The effect of a complete cutoff is shown most strikingly when one of the curves is rectangular, as in Figs. 6D and 6E. This is the sharpest cutoff conceivable, and the other curve of the pair acquires negative and positive spurious pulses trailing off gradually either side of the desired center pulse.[10] The widening of this curve, is evidenced by an increase of the nominal width of the center pulse to 1.18 times that of the entire curve, while the total absolute area of all pulses becomes very great (theoretically infinite) because they decay so slowly either side of the center.

The cosine-squared curve of Fig. 6F offers an attractive compromise, with an absolute cutoff beyond twice its nominal width. Its cutoff is so gradual that the resulting widening of its companion curve is negligible (about 1 per cent), as are also the spurious pulses either side of the center pulse. In fact, its companion curve is almost exactly the

[10] The widening of the central pulse and the addition of spurious pulses in one curve, caused by complete cutoff in the other curve, are precisely analogous to the diffraction patterns familiar in optics.

same shape, and may be regarded as the same for practical purposes.

In Fig. 6, the left-hand curve $G(t)$ of each pair represents a transient signal which may be obtained, at least theoretically, by a certain shape of scanning spot or aperture in a transmitter, scanning a bright and narrow vertical line object. The right-hand curve $F(f)$ represents the frequency spectrum of the signal, which should be passed by the electrical filters if distortion is to be avoided. In Fig. 6A, the signal is obtained from a very narrow aperture; in Figs. 6B and 6C, from a scanning spot obtained by an electron beam of Maxwellian distribution of velocities;[11] in Fig. 6D, from a rectangular aperture; and in Fig. 6F, from a circular scanning spot of the derived ideal distribution.

The width of a scanning spot or aperture is seen to cause the attenuation of the higher-frequency components in the frequency spectrum. Therefore an aperture of finite width, as compared with a very narrow aperture, is equivalent to the introduction of an electrical filter which would have the same effect.[12] A very narrow aperture scanning a narrow vertical line yields the signal impulse $G(t)$ and the uniform frequency spectrum $F(f)$ of Fig. 6A. Therefore the wider spots or apertures are equivalent to inserting electrical filters of characteristics such as $F(f)$ in the other examples of Fig. 6.

The scanning spot or aperture of finite width in the receiver is the equivalent of introducing another filter of the characteristics identified with the shape and size of this aperture. In Fig. 6G, there is shown the combined effect of the two ideal scanning spots in transmitter and receiver. The curve $G(t)$ is the resultant horizontal distribution of a vertical line image. It does not exist as a time variation of the signal, because the second spot is beyond all the signal circuits. It exists only as a horizontal variation of intensity on the screen in the receiver. Likewise the frequency spectrum $F(f)$ does not exist electrically. It represents rather the filter which, if interposed between scanning apertures of negligible width, would give the final image the horizontal distribution shown by the transient curve $G(t)$.[13] This filter is the combination of two filters according to Fig. 6F, one resulting from the width of each of the scanning spots. Its nominal cutoff frequency is $\frac{3}{4}$ that of each of the component filters, corresponding to the 4/3 increase of the horizontal width, derived in Appendix C.

Having the resultant frequency characteristics of both scanning systems, there must be stated also a desired over-all frequency characteristic of the system as a whole, so the required electrical characteris-

[11] Maloff and Epstein (14).

[12] Gray, Horton, and Mathes (5).

[13] This transient curve is derived in Appendix C, while the frequency curve is obtained by squaring the ordinates of the curve just above.

tics can be specified. The electrical filters of the system must cut off completely above a certain frequency, or beyond a certain band width, because only a limited part of the frequency spectrum is allotted to a single television transmitter. Also the specified filter cutoff must be sufficiently gradual to be obtainable in practice. The system as a whole must have a gradual cutoff for another reason, to avoid causing appreciable spurious pulses in the transient signal, or rather in the final image.

The logical horizontal distribution to be desired across a vertical line image would be the same as the vertical distribution across a nearly horizontal line image, but this is not definite because it is somewhat variable along the line as shown in Fig. 3. Its average width is definite, however, and is $\sqrt{2}$ times the line pitch. Its average distribution is approximately that of a cosine-squared curve of this width, so this simple form is chosen as the desired horizontal distribution across a vertical line image. In other words, this is the desired over-all curve of $G(t)$ for the entire system. This cannot be used practically because it cannot be obtained with a complete cutoff in the low-pass filters. Nearly the same shape can be obtained, however, by interchanging the two similar curves in Fig. 6F, so the cosine-squared curve with its complete cutoff appears in the frequency column. Therefore this shape of frequency characteristic is chosen as a practical and nearly ideal over-all characteristic for the entire system.[14] The transient curve of Fig. 6F, both before and after the interchanging, has a width equal to that of the scanning line, which is the line pitch. The width of the over-all frequency curve should be less in the ratio $1/\sqrt{2}$, to expand the horizontal width of the transient curve to $\sqrt{2}$ times the line pitch, making equal the horizontal and vertical width of confusion. This makes the desired over-all frequency curve nearly the same as the curve of Fig. 6G for the two scanning spots, but about 6 per cent narrower.

Fig. 7 shows the frequency characteristics of the system as a whole. The over-all nominal cutoff frequency f_c is that at which one-half period corresponds to the horizontal width of confusion or $\sqrt{2}$ times the line pitch. The lower solid curve is the desired over-all characteristic, a cosine-squared curve having nominal cutoff at f_c and complete cutoff at $2f_c$ (Fig. 6F, interchanged). The upper solid curve is the frequency characteristic of both scanning spots (Fig. 6G). The cutoff frequency of one spot on this scale is $\sqrt{2}\ f_c$ so that of both spots is $(3/4)\sqrt{2}\ f_c = 1.06 f_c = f_{cs}$.

[14] The choice of the cosine-squared curve here is not unique, but is logical for consistency with the unique choice of that curve for the transverse distribution of scanning lines, where the requirements are definite.

The ratio of the two solid curves in Fig. 7, shown by the heavy dashed curve, is the required frequency characteristic of all low-pass filters taken together, if the chosen conditions are to be realized. It is nearly flat to f_c and has a gradual cutoff between f_c and $2f_c$. On the basis of the area of this curve, the nominal cutoff frequency of the electrical filters is $1.62f_c = f_{ce}$. The shape of the filter curve is not critical in the region of gradual cutoff, because the frequency components are relatively weak. The light dotted curves above and below the filter curve give some idea of the relative tolerance of departure from the curve shown. Assuming a tolerance of 5 per cent either way at zero frequency, it is 10 per cent at f_c and 100 per cent at $1.7f_c$. A sharper cutoff slightly below $2f_c$ is tolerable from the viewpoint of the amount of distortion it

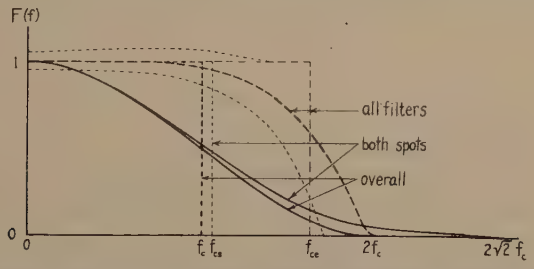


Fig. 7—The over-all frequency characteristics obtained by the transmitter and receiver scanning spots and the electrical low-pass filters.

would cause, but it is difficult to obtain in practice. Therefore $2f_c$ is theoretically and practically the most logical frequency of complete cutoff of the entire filter system of both transmitter and receiver.

Disregarding practical considerations of filter design and disregarding the detrimental effect of spurious pulses on either side of the center pulse, it is instructive to investigate what is the frequency of sharp cutoff that would give the same desired area of the center pulse. If the over-all curve were rectangular in shape (obtained by filters compensating for the gradual cutoff of the scanning spots) this area would be obtained with a sharp cutoff at $1.18f_c$. It is concluded that a practical system giving even this compromised performance would have complete cutoff[15] at a frequency considerably above $1.18f_c$.

In concluding the discussion of the horizontal resolution and the applications of the Fourier integral, there is proposed a useful rule for estimating approximately the width of a line image in terms of the contributions made by various components of the system. This rule is

[15] Mertz and Gray (12) give a different logical basis of specifying curves of different shape and equivalent width. Referring to their Fig. 12 (A and E) as applied to the present discussion, a rectangular cutoff at $1.25f_c$ would be equivalent to the cosine-squared cutoff at $2f_c$.

based on the example of the probability curves, but is approximate for most cases where the cutoff is gradual. First the nominal widths of the transient curves are computed individually for the respective components, such as the object line, optical systems, screens (grain size), scanning spots, and filters. Then the resultant width of the image line, as derived in Appendix D for probability curves, is the quadratic sum of all components

$$w = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}. \tag{3}$$

This is termed the "quadratic-sum rule." The degree of approximation in practical cases is greatest where the individual frequency characteristics are close to the shape of probability curves (which is true of the cosine-squared curve), and is least where they are flat and have sharp cutoff.

IV. The Requirements of a Television Frequency Channel

The frequency requirements of a television signaling system are important from the standpoint of both engineering design and the allocation of frequency channels, so this phase of the problem has received much attention. The least definite factor in the previous formulas has been that expressing the condition for equal horizontal and vertical resolution, which is treated quite fundamentally in this paper. Therefore it is now possible to derive the desired formula on a more definite theoretical basis, especially as regards the real significance of the nominal cutoff frequency of the entire system. In order to simplify the formula, the loss of picture area during retrace intervals is not taken into account.

For a given number of scanning lines N_s, the height of the picture image is N_s times the pitch of scanning lines, or $N_s/\sqrt{2}$ times the vertical width of confusion. The horizontal width of confusion is equal to the vertical. Therefore, for an aspect ratio R, the width of the picture is $RN_s/\sqrt{2}$ times the horizontal width of confusion. In other words, a single scanning line contains $RN_s/\sqrt{2}$ half periods or $RN_s/2\sqrt{2}$ periods at the nominal cutoff frequency f_c. Multiplying this number of periods by the number of lines, N_s, and by the frame frequency f_p, the nominal cutoff frequency is found to be

$$f_c = f_p N_s^2 R/2\sqrt{2}. \tag{4}$$

The factor $1/\sqrt{2} = 0.71$ has been determined experimentally to have values in the range of 0.64 to 0.75, so the agreement is very close.[16] If

[16] Kell, Bedford, and Trainer (13) and (20).

a picture element is defined as a square of the width of confusion, the total number of elements is $N_s^2R/2$.

An important result of the present treatment is the specification of over-all filter characteristics which are proved to be adequate, and which are not critical so they can be realized with sufficiently close approximation in practice.

A television signal, according to the present proposals, comprises both a picture signal and a sound signal included in an allotted frequency channel. The most favorable arrangement within the channel is probably along the lines of Fig. 8. The sound carrier is spaced from the

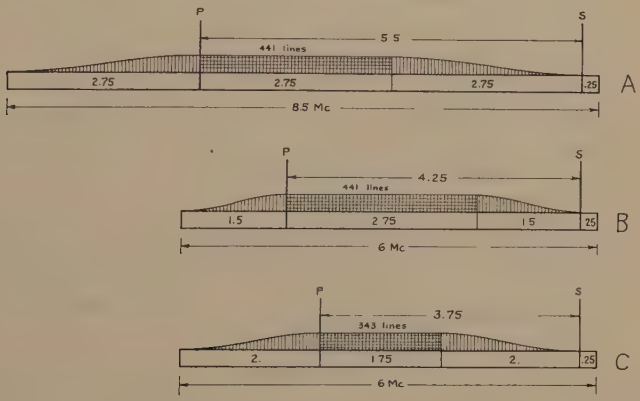


Fig. 8—The subdivision of a frequency channel to accommodate one complete side band of the picture signal.

upper edge of the channel by a guard band of 0.25 megacycles (about 1/2 per cent, if the channel is in the vicinity of 50 megacycles). The picture carrier P is located in the lower part of the channel, with the intent of transmitting only the upper side band in its entirety, along with the inner part of the lower side band. This provides double-side-band transmission of only the stronger modulation components of lower frequencies, for which distortionless detection requires both side bands.

Gradual cutoff of the partial lower side band is necessary also to insure sufficient tolerance of filter design and of mistuning in the receiver. Gradual cutoff of the complete upper side band is desirable beyond the nominal cutoff frequency, for the same reasons and also to preserve the horizontal resolution. It is about equally important on both sides of the nominal single side band which is transmitted completely and carries most of the information.

Fig. 8A shows the subdivision of a frequency channel for full utilization of 30 frames per second, 441 lines, and 4/3 aspect ratio.

The nominal width of the single side band is $f_c = 2.75$ megacycles. The carriers P and S are separated by $2f_c$, the frequency of complete cutoff, so the gradual cutoff toward the sound carrier occupies an added width equal to f_c. An equal width is allotted to the gradual cutoff of the partial side band on the other side of the carrier. The outer edges of the gradual cutoff bands are sufficient as guard bands on the outer edges of the picture signal. On this basis, the required width of the frequency channel is 8.5 megacycles.

If the channel width is limited to 6 megacycles, Fig. 8B shows a fairly good compromise obtained by reducing from 2.75 to 1.5 megacycles, the width of each gradual cutoff. This leaves about the sharpest cutoff realizable without excessively impairing the equality of horizontal and vertical resolution.

On the other hand, Fig. 8C shows how a channel width of 6 megacycles is more than adequate for 343 lines, for which the nominal cutoff frequency is only 1.66 megacycles (1.75 being provided in the diagram).

It is concluded that 30 frames per second and 441 lines, which ideally requires 5.5-megacycle separation between picture and sound carriers, can barely be accommodated in a channel 6 megacycles wide, with 4.25-megacycle carrier separation.

V. Conclusion

The following relations in a television system are proved on reasonable grounds:

(1) A flat field of scanning in the transmitter is required to minimize spurious large patterns caused by long beads on the images of nearly horzontal lines.

(2) A flat field of scanning in the receiver is required to obtain a flat image of a flat object, that is, an object of uniform illumination.

(3) A flat field at either end is best obtained by scanning lines of cosine-squared transverse distribution, generated by a circular spot designed to secure this distribution.

(4) The resulting vertical width of confusion is $\sqrt{2}$ times the pitch of scanning lines.

(5) The two scanning spots together have a horizontal width of confusion 4/3 times the line pitch.

(6) Equal horizontal and vertical width of confusion requires all the electrical filters of the system to pass uniformly the components out to the nominal cutoff frequency f_c, at which one-half period corresponds to the width of confusion.

(7) Practical filter design limitations and the avoidance of spurious

pulses in the image require that the filters have a gradual cutoff between f_c and $2f_c$, while they may have complete cutoff beyond $2f_c$.

Acknowledgment

The authors gratefully acknowledge the assistance of their associates and particularly the helpful suggestions and criticism of Messrs. D. E. Harnett, H. M. Lewis, J. C. Wilson, and M. Cawein.

Appendix A

The Vertical Width of Confusion

The vertical width of confusion is the average width of the image of a narrow bright line nearly horizontal and therefore nearly parallel with the scanning lines. Fig. 4 shows the basis for defining the effective width of a line image at a point on the line, in terms of the transverse distribution of light intensity. Fig. 3 shows as a basis the varying relative contributions of two adjacent scanning lines in the reproduction of such a line.

The effective width of the reproduced line is defined as the width of a rectangle having the same area, whose height is equal to the ordinate at the point where the line would cross if the reproduction were perfect. This ordinate is v in Fig. 4, where the centers of the two adjacent scanning lines are at $x=0$ and 1, and the sloping line to be reproduced is at u times the distance from one scanning line ($x=0$) to the next ($x=1$). The first step is to determine v in terms of u, for values of u between 0 and 1. The relative intensities of the two lines are, respectively,

$$v_0 = \frac{1 + \cos \pi u}{2}; \qquad v_1 = \frac{1 + \cos \pi(1 - u)}{2} = \frac{1 - \cos \pi u}{2}. \tag{5}$$

The total intensity of the two lines and therefore the total area is unity, which is in accordance with the conditions placed on the scanning lines at the transmitter. The total ordinate at any point is

$$\begin{aligned} v &= v_0 \frac{1 + \cos \pi u}{2} + v_1 \frac{1 + \cos \pi(1 - u)}{2} \\ &= \frac{(1 + \cos \pi u)^2}{4} + \frac{(1 - \cos \pi u)^2}{4} = \frac{1 + \cos^2 \pi u}{2}. \end{aligned} \tag{6}$$

Since the area is unity and the ordinate is v, the effective width of reproduction of the line is

$$w = \frac{1}{v} = \frac{2}{1 + \cos^2 \pi u}. \tag{7}$$

The minimum width is one, the maximum is two, and the average width is

$$w_a = \int_0^1 w du = \sqrt{2}. \tag{8}$$

This is a special case in which the "quadratic-sum rule" is exactly applicable to the average combined width of scanning lines in transmitter and receiver. The resulting vertical width of confusion is $\sqrt{2}$ times the pitch of scanning lines.

Appendix B

The Circular Spot Required to Give the Ideal Scanning Line

The scanning line is best obtained by means of a circular spot with such distribution of light as to give the ideal transverse distribution in the line, as shown in Fig. 5.

Across the line, the ideal distribution is

$$v = \cos^2 \frac{\pi}{2} u = \frac{1 + \cos \pi u}{2} \tag{9}$$

in which u is the relative distance from the center of one line to the center of the next. For purposes of integration, it is necessary to write this relation in algebraic form. An infinite series could be used, but an excellent approximation which is better suited to present needs is

$$v' = a(1 - u^2)^2 + b(1 - u^2)^4; \quad -1 < u < 1 \tag{10}$$

in which a and b are constant coefficients so chosen that $v' = v = 1$ at $u = 0$, and that the areas are equal:

$$\int_{-1}^{+1} v' \cdot du = \int_{-1}^{+1} v \cdot du = 1 \tag{11}$$

so the curves must coincide at one or more intermediate points on each side. By the choice of this form of approximation, the curves are both level at $u = 0$ and ± 1, and also they coincide at $u = \pm 1$. The coefficients, evaluated to meet these conditions, are

$$a = \frac{59}{80} = 0.738; \qquad b = \frac{21}{80} = 0.262. \tag{12}$$

Fig. 9 shows one quadrant of a circular spot, whose required radial distribution is to be determined.[17] The x direction is across the line.

[17] Gray, Horton, and Mathes (5), Fig. 30.

The y direction is along the line, the direction of motion of the spot. The z ordinates represent relative intensity. The relative intensity across the line is the relative area of a cross section in the plane, $x = u$. Therefore the shape of the distribution must be determined from the variation of this area.

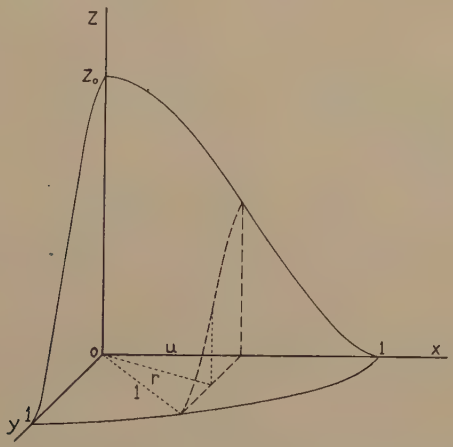


Fig. 9—Derivation of the circular distribution of the scanning spot.

The circular distribution requires that the ordinates depend only on the radius:

$$z = f(r); \qquad r = \sqrt{x^2 + y^2}. \tag{13}$$

The following radial function z' is chosen as the algebraic approximation for z, because it yields the expression for v' on integration

$$\begin{aligned} z' &= a'(1 - r^2)^{3/2} + b'(1 - r^2)^{7/2}; \qquad r < 1 \\ &= a'(1 - u^2 - y^2)^{3/2} + b'(1 - u^2 - y^2)^{7/2} \end{aligned} \tag{14}$$

in which a' and b' are constant coefficients to be evaluated.

The area of the cross section in two quadrants is

$$v' = 2\int_0^{\sqrt{1-u^2}} z' \cdot dy. \tag{15}$$

For integration, let

$$\begin{aligned} \frac{y}{\sqrt{1 - u^2}} &\equiv \sin q; \qquad \frac{\sqrt{1 - u^2 - y^2}}{\sqrt{1 - u^2}} = \cos q; \\ \frac{dy}{\sqrt{1 - u^2}} &= \cos q \cdot dq \end{aligned} \tag{16}$$

$$v' = 2a'(1-u^2)^2 \int_0^{\pi/2} \cos^4 q \cdot dq + 2b'(1-u^2)^4 \int_0^{\pi/2} \cos^8 q \cdot dq \quad (17)$$

$$= \frac{3\pi}{8} a'(1-u^2)^2 + \frac{35\pi}{128} b'(1-u^2)^4.$$

The coefficients are now readily evaluated:

$$a' = \frac{8}{3\pi} a = \frac{59}{30\pi} = 0.626; \qquad b' = \frac{128}{35\pi} b = \frac{24}{25\pi} = 0.302. \quad (18)$$

The center ordinate is

$$z_0' = a' + b' = 0.928 \quad (19)$$

from which the coefficients relative to unity at the center are

$$a'/z_0' = 0.675; \qquad b'/z_0' = 0.325. \quad (20)$$

It is interesting to determine also the effective diameter of the ideal spot, which is defined as that of a cylinder of equal volume and height. The volume of the ideal spot is

$$V = \int_0^1 z(2\pi r)dr. \quad (21)$$

For integration, let

$$s \equiv \sqrt{1-r^2}; \qquad r = \sqrt{1-s^2}; \qquad r \cdot dr = -s \cdot ds. \quad (22)$$

Then the volume of the approximation is

$$V' = 2\pi a' \int_0^1 s^4 ds + 2\pi b' \int_0^1 s^3 ds \quad (23)$$

$$= \frac{2\pi}{5} a' + \frac{2\pi}{9} b' = \frac{59}{75} + \frac{16}{75} = 1.$$

This unit volume is a check on the previous analysis, since it corresponds to the condition of unit area under the approximate curve v'. The volume, in terms of the effective radius r_e is

$$V = \pi r_e^2 z_0 \quad (24)$$

so the effective diameter is

$$2r_e = \sqrt{4V/\pi z_0} = 1.17 \quad (25)$$

as compared with unit spacing between scanning lines, or as compared with a 2-unit outside diameter of the spot.

When the ideal spot moves horizontally in a scanning line, the line generated is uniform along its length, but when the same spot is used to synthesize a vertical line by horizontal scanning, the resulting vertical line comprises a succession of individual spots, with adjacent spots overlapping. Such an image would be produced in the receiver if the transmitter scanning spot had no width. It is desired to verify the average width of such a line image, which should be the same as that of a scanning line.

The cross section of this vertical line image has unit area at all points along the line, as a result of the flat field of scanning. The total intensity along the center of the line has a minimum value at the center of each spot and a maximum halfway between centers. From (14), (18), and (19), the minimum value at $r=0$ is $z_0'=0.928$, and the maximum value is the sum of two spots at $r=1/2$, namely, $2z'=1.032$. The average intensity along the center of the line image is unity, because there is one spot per unit of length (equal to the line pitch), and the intensity area of each spot is unity in its central cross section (one of the conditions in the evaluation of a' and b'). The width is then the reciprocal of the peak intensity along the center line, and varies between 0.97 and 1.08 times the pitch of scanning lines. Since the intensity is so near unity at all points, and its average value is unity, the average width is extremely close to unity. It is concluded that the vertical succession of ideal circular scanning spots forms a line nearly the same as the horizontal scanning line formed by the motion of the same spot.

Appendix C

The Horizontal Width of Confusion of Two Circular Scanning Spots

There is to be derived the horizontal width of confusion of two circular scanning spots, each having the ideal characteristics required to give the cosine-squared lateral distribution of the scanning line. This is the width of the vertical line image secured by scanning a vertical narrow bright line as the object.[18] This image is not precisely uniform along its length, because it is composed of individual spots displaced vertically by the pitch of the scanning lines. However, these spots overlap sufficiently to secure nearly uniform width of the vertical line image.

Fig. 10 shows the horizontal characteristics of a narrow vertical line object scanned horizontally in the transmitter and receiver by vertical lines, each representing a vertical succession of ideal scanning spots. In other words, each vertical array of scanning spots is conceived to be replaced by a vertical line (here denoted a "line-spot") of

[18] Gray, Horton, and Mathes (5), Fig. 29.

the same cosine-squared lateral distribution as the horizontal scanning line. One of these line-spots moves broadside across the line object in the transmitter, causing a flare-up of the other line-spot in the receiver as it crosses the place where the line image should appear. Obviously, the extreme width of the line image is double that of the scanning spot, or four times the scanning-line pitch. The effective width, however, is much less, and is here to be derived. (This problem is analogous to the derivation of the average width of the nearly horizontal line image, of which the extreme width is three times the pitch, but the average width is only $\sqrt{2}$ times the pitch.)

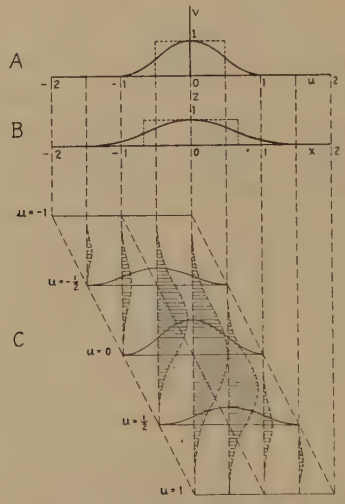


Fig. 10—Derivation of the horizontal width of a vertical line image.

Instead of moving the line-spot in the transmitter, it is held stationary and the object line is moved, which gives the same result. Fig. 10A shows the cosine-squared lateral distribution of the stationary line-spot in the transmitter. In other words, v is the relative response for a horizontal distance u between the moving narrow line and the center of the line-spot. Fig. 10B shows the lateral distribution of the resulting vertical line image in the receiver, which is obtained by the integrations indicated in Fig. 10C.

In Fig. 10C, the respective horizontal curves at different levels represent the intensity of the receiver line-spot in successive positions as it crosses the location of the image. The distance between the centers of the receiver line-spot and the image is u, the same as between the centers of the transmitter line-spot and the object. Therefore the respective levels correspond to successive values of u. The intensity

of the image at a fixed distance x from its center is the total contributed by the receiver line-spot in all positions ($-1<u<1$). The contribution in each position is denoted y and is plotted sideways to define a vertical shaded area at each distance x, this area being the total contribution of the receiver line-spot in all positions. (The vertical scale of these areas is doubled to reduce the confusion of curves.) The shaded area is denoted z as a function of the distance x from the center of the image, and is plotted in Fig. 10B as the horizontal distribution across the vertical line image.

In Fig. 10A, the relative response, at a horizontal distance u between line object and line-spot, is

$$v = \cos^2 \frac{\pi}{2} u; \qquad -1 < u < 1. \tag{26}$$

The relative contribution of the receiving line-spot, in each phase, to the image intensity, is

$$y = v \cos^2 \frac{\pi}{2}(x-u) = \cos^2 \frac{\pi}{2} u \cdot \cos^2 \frac{\pi}{2}(x-u). \tag{27}$$

The total area of contribution, at a distance x from the center, is

$$z = \int_{-1}^{x+1} y \cdot du \qquad \text{for} \quad -2 < x < 0 \tag{28}$$

$$z = \int_{x-1}^{+1} y \cdot du \qquad \text{for} \quad 0 < x < 2. \tag{29}$$

These expressions are valid on the respective sides of the center of the image, but the image is symmetrical, so both yield the same shape relative to the center. Noting that

$$y = \left[\cos \frac{\pi}{2} u \cdot \cos \frac{\pi}{2}(x-u)\right]^2 \tag{30}$$

$$= \frac{1}{4} \cos^2 \frac{\pi}{2} x + \frac{1}{4} \cos^2 \frac{\pi}{2}(2u-x) + \frac{1}{2} \cos \frac{\pi}{2} x \cdot \cos \frac{\pi}{2}(2u-x)$$

and integrating,

$$\int y \cdot du = \frac{u}{4} \cos^2 \frac{\pi}{2} x + \frac{2u-x}{16} + \frac{1}{16\pi} \sin \pi(2u-x) + \frac{1}{2\pi} \cos \frac{\pi}{2} x \cdot \sin \frac{\pi}{2}(2u-x). \tag{31}$$

Substituting the limits of u, for negative values of x,

$$z = \frac{2+x}{8}(2+\cos \pi x) - \frac{3}{8\pi}\sin \pi x \tag{32}$$

and for positive values of x,

$$z = \frac{2-x}{8}(2+\cos \pi x) + \frac{3}{8\pi}\sin \pi x. \tag{33}$$

The center ordinate is $z_0 = 3/4$, and the area under the z curve is

$$s = \int_{-2}^{+2} z \cdot dx = 1 \tag{34}$$

so the width of the line image is $w = s/z_0 = 4/3$ relative to unit pitch of the scanning lines. This is the resultant horizontal width of confusion, and is slightly less than $\sqrt{2}$, the vertical width of confusion with the same circular scanning spots.

Appendix D

The Relation Between the Frequency Spectrum and the Horizontal Width of Confusion

The horizontal width of confusion is the width of the image of a very narrow vertical line object. While it is explicitly a space dimension, it is implicitly dependent on the velocity of the scanning spot and the variation of its brightness with time. The correlation of this time function with the frequency characteristics of the apparatus is obtained by the use of the Fourier integral, the limiting case of the more familiar Fourier series. The general application of the Fourier integral to such a problem is complicated because it involves the frequency characteristics of both amplitude and phase.

In the present case, however, phase distortion is not an essential to the problem because it can be canceled to any required degree of approximation by means of phase-correcting networks. The only essential limitation is the width of the frequency spectrum. Therefore, the condition of negligible phase distortion is here assumed; in fact, zero phase shift is assumed, because the slope of the linear phase shift, remaining after phase correction, causes only a negligible delay of the reproduction of the entire picture. This removal of the phase shift from the problem greatly simplifies the use of the Fourier integral without detracting from the generality of the conclusions to be drawn.[19]

[19] This simplification has been employed by Starr (15) and Guillemin (16) in the analysis of transients in ideal filters. Also it has appeared in the analysis of the frequency spectrum of symmetrical transients or apertures, by Gray, Horton, and Mathes (5), Mertz and Gray (12), and Wilson (18).

In this treatment, the general forms of Campbell and Foster are followed in order to facilitate further reference to their works.[20] In the absence of phase shift, any amplitude spectrum corresponds to a transient symmetrical on the time axis, which yields a symmetrical picture element in the scanning process.

The simplified forms of the Fourier integral, with zero phase shift, are

$$F(f) = \int_{-\infty}^{+\infty} G(t) \cos 2\pi ft \cdot dt = 2 \int_{0}^{\infty} G(t) \cos 2\pi ft \cdot dt \qquad (35)$$

$$G(t) = \int_{-\infty}^{+\infty} F(f) \cos 2\pi ft \cdot df = 2 \int_{0}^{\infty} F(f) \cos 2\pi ft \cdot df \qquad (36)$$

in which $F(f)$ is the amplitude-frequency spectrum and $G(t)$ is the time variation of the transient. The second expression of each function results from the symmetry of the first. The amplitude-frequency spectrum $F(f)$ is fundamentally symmetrical about zero frequency, and the cosine is also symmetrical, so the time variation $G(t)$ is symmetrical.

The present treatment deals with low-pass filters restricting the band width of the frequency spectrum, so $F(f)$ is taken as unity at zero frequency. Then this function represents the relative-amplitude charteristic of the system. If this characteristic were obtained by a simple filter, the quantity of electrical input and output would be the same. The more general significance of this relation is equality of areas under the amplitude-time curves $G(t)$ expressing the respective input and output transients. The significance in scanning is the equality of the amount of light in a line object at the camera tube and a line image at the picture tube, regardless of the width of confusion covered by the light.

There are next defined two factors expressing the width of the frequency band and the time width of the transient, which corresponds to the space width of confusion in the picture. Letting $F_0 = 1$ be the value of $F(f)$ at $f = 0$, as mentioned above, and letting G_0 be the amplitude of the transient $G(t)$ at the center, $t = 0$,

$$F_0 = \int_{-\infty}^{+\infty} G(t) \cdot dt \equiv G_0(2t_c) = 1 \qquad (37)$$

$$G_0 = \int_{-\infty}^{+\infty} F(f) \cdot df \equiv F_0(2f_c) = 2f_c. \qquad (38)$$

[20] Campbell and Foster (8).

These relations serve to define the nominal width of the transient $2t_c$ and the nominal cutoff frequency f_c of the system acting as a low-pass filter. Each is based on the width of a rectangle having the same center ordinate and the same area as the corresponding curve. (The area of the curve is the net area, or the difference of the areas above and below the axis.) The preceding equations, taken together, give the relation between the band width and the transient width,

$$(2f_c)(2t_c) = 1. \tag{39}$$

The transient width is one-half period at the cutoff frequency. This is a well-known relation but its exact basis has not been generally appreciated.

In the usual treatment of the Fourier integral, $F(f)$ is described as the frequency spectrum of a transient $G(t)$. The alternative concept is also required here. The transient $G(t)$ is the output of a low-pass filter $F(f)$ in response to a so-called "unit impulse," which has unit area obtained by very large amplitude and very small width. Since $F_0 = 1$, the output transient also has unit area, as indicated above. Therefore its center amplitude is equal to the reciprocal of its width.

These concepts together form the basis for the correlation of electrical-filter distortion with such distortion as is caused by space dimensions of the scanning spot or aperture, and of optical dispersion. If the space distortion in scanning causes a transient $G(t)$, that is equivalent to inserting in the electrical circuit a filter $F(f)$. The characteristics of filters are easily multiplied together, so the total distortion from all causes is easily expressed, and the corresponding transient is derived by means of the Fourier integral. This transient is the output of the entire system in response to an instantaneous impulse, which corresponds to scanning a very narrow vertical line.

The unique case of the Fourier integral is that in which the frequency and time functions are the same:[21]

$$F(f) = \exp(-\pi f^2); \qquad G(t) = \exp(-\pi t^2). \tag{40}$$

These functions are well known as the "probability curve." They are shown in Fig. 6C. Each has unit height and unit area, that is, $F_0 = G_0 = 1$ and $2f_c = 2t_c = 1$. These may be modified in relative dimensions while retaining the same form of curve:[22]

$$F(f) = \exp(-\pi f^2/4f_c^2) = \exp(-4\pi f^2 t_c^2) \tag{41}$$

[21] Campbell and Foster (8), No. 705.1 in the tables.
[22] Campbell and Foster (8), No. 708.0 in the tables.

$$G(t) = \frac{1}{2t_c} \exp\,(-\pi t^2/4t_c^2) = 2f_c \exp\,(-4\pi f_c^2 t^2). \qquad (42)$$

The transient still has unit area, but increasing the cutoff frequency causes its width to decrease and its height to increase. As the cutoff frequency increases, filter distortion becomes less and the transient approaches the unit impulse in its dimensions. An intermediate case and the extreme case are shown in Figs. 6B and 6A. (A Maxwellian distribution of electrons in a scanning beam would yield this type of transient.)

This unique case forms the basis for the "quadratic-sum rule," which is proposed for computing the resultant transient width of a number of devices in terms of their individual characteristics. Substituting $w = 2t_c = \frac{1}{2}f_c$ for the transient width,

$$F_1 = \exp\,(-\pi f^2 w_1^2), \quad F_2 = \exp\,(-\pi f^2 w_2^2), \quad \text{etc.} \qquad (43)$$

$$F = F_1 F_2 \cdots F_n = \exp\,[-\pi f^2 (w_1^2 + w_2^2 + \cdots + w_n^2)] \qquad (44)$$

$$= \exp\,(-\pi f^2 w^2)$$

in which the resultant band width is the quadratic sum of the partial band widths,

$$w = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}. \qquad (45)$$

This rule is exact only for this case but is a useful approximation in practical cases where the shape of transients and filter characteristics is not too sharp and therefore does not depart too much from the shape of the probability curve.

As an introduction to the practical application of these principles, there is the case of a rectangular aperture scanning a narrow vertical line in a time of $2t_c$. This is shown in Fig. 6D. The frequency spectrum of the resulting rectangular transient, or the equivalent filter characteristic, is obtained by the Fourier integral, letting $G(t) = 1/2t_c$ in the interval $(-t_c < t < +t_c)$:

$$F(f) = 2\int_0^{t_c} \frac{1}{2t_c} \cos 2\pi f t \cdot dt = \frac{\sin 2\pi f t_c}{2\pi f t_c} = \frac{\sin \pi f/2f_c}{\pi f/2f_c}. \qquad (46)$$

This frequency spectrum decays slowly in amplitude toward higher frequencies, so a wide frequency band would be required for its reproduction. The required frequency band is much wider than f_c because the total of positive and negative areas is much greater than the net area.

The converse of this case has a rectangular filter characteristic, as shown[23] in Fig. 6E. Assuming $F(f)=1$ in the interval $(-f_c<f<+f_c)$, the resulting transient is

$$G(t) = 2\int_0^{f_c} \cos 2\pi ft \cdot df = 2f_c \frac{\sin 2\pi f_c t}{2\pi f_c t} . \qquad (47)$$

This is the transient response of this idealized filter to a unit impulse.

This example is given to show the correspondence between the frequency spectrum for a given transient, on one hand, and the transient response of a given filter to an impulse, on the other hand. The correspondence is an outgrowth of the uniform frequency spectrum of the instantaneous impulse. If the input to the filter has any other transient form, and therefore a nonuniform frequency spectrum, that is the same as applying an impulse through an additional "shaping" filter which modifies its frequency spectrum.[24] The output transient then has a frequency spectrum determined entirely by the resultant characteristics of the "shaping" filter and the actual filter, and its shape may be computed accordingly by the Fourier integral. The frequency spectrum of the input transient is merely conceived as the characteristic of the additional filter contributing to the distortion of the assumed input, which is an impulse.

The idealized rectangular filter characteristic yields, as shown in Fig. 6E, a transient which is very wide in its influence, even though its theoretical width is only $2t_c$ based on its net area. The center pulse by itself has more than the net area, its effective width being 2.36 t_c. Then there is also the slowly (hyperbolically) damped oscillation whose frequency is the cutoff frequency f_c. In scanning a bright vertical line, the spurious positive pulses appear as multiple images and the negative pulses as shadows. The effect is analogous to the diffraction patterns of optics.

It appears that a sharp cutoff in the frequency characteristic yields a transient of such width as not to be adapted for television scanning. Therefore the system as a whole must have effectively a gradual cutoff which requires an actual frequency band wider than the nominal width f_c.

The opposite extreme is a filter characteristic shaped like the probability curve, as in Fig. 6C. This seems to have the sharpest cutoff possible without causing discrete spurious pulses in the transient response to an impulse. The attenuation curve on logarithmic ordinates is a parabola. Theoretically, this characteristic requires components of all

[23] Starr (15), Guillemin (16).
[24] Nyquist (3).

frequencies, but practically those of frequencies beyond $2f_c$ are negligible (less than 4.3 per cent in amplitude, and less than 1.2 per cent in area). Likewise, the transient response is negligible beyond $2t_c$, and the spurious pulses occasioned by neglecting the higher-frequency components are small.

The preceding extreme cases suggest a compromise filter characteristic having a gradual cutoff but no transmission above a complete cutoff at $2f_c$, twice the nominal cutoff frequency. The shape of the characteristic which gives the most gradual transition between the required values of unity at zero frequency and zero at $2f_c$ is the cosine-squared shape (like the transient of Fig. 6F), since this is tangent to the horizontal at the limits. It also meets the requirement that the area must equal $2f_c$.

$$F(f) = \cos^2 \pi f/4f_c = \cos^2 \pi f t_c; \qquad -2f_c < f < 2f_c \tag{48}$$
$$= \frac{1 + \cos \pi f/2f_c}{2} = \frac{1 + \cos 2\pi f t_c}{2}.$$

The output transient in response to an impulse then has the form

$$G(t) = 2\int_0^{2f_c} \frac{1 + \cos \pi f/2f_c}{2} \cos 2\pi f t \cdot df \tag{49}$$
$$= 2f_c \frac{\sin 4\pi f_c t}{4\pi f_c t} + f_c \frac{\sin (\pi - 4\pi f_c t)}{\pi - 4\pi f_c t} + f_c \frac{\sin (\pi + 4\pi f_c t)}{\pi + 4\pi f_c t}$$
$$= \frac{2f_c}{1 - (4f_c t)^2} \cdot \frac{\sin 4\pi f_c t}{4\pi f_c t}.$$

The intermediate expression is useful for securing indeterminate values. When any denominator is zero, that fraction reduces to unity and the other terms reduce to zero. The last expression shows the rapid (third-power) decay of ordinates outside the nominal width of the transient, which is between $-t_c$ and $+t_c$. This case is Fig. 6F with the time and frequency curves interchanged.

The shape of this transient very closely approximates the shape of the filter characteristic, which is, in the terminology of the transient,

$$G(t) = 2f_c \cos^2 \pi f_c t = (1/2t_c) \cos^2 \pi t/4t_c; \qquad -2t_c < f < +2t_c. \tag{50}$$

The respective curves, drawn for $2f_c = 1$, have in common the unit ordinate at the center, the half-unit ordinates at $\pm t_c$, and zero ordinates at integral multiples of $\pm t_c$. The exact transient curve has a maximum negative value less than 3 per cent. The center pulses agree

in area within less than 1 per cent. Therefore, in this case, the shape of the filter characteristic and the shape of the transient having the same spectrum may be regarded as the same for practical purposes. In other cases, this relation cannot be obtained so closely if the absolute cutoff occurs at less double the nominal cutoff frequency.

Bibliography

(1) Lord Rayleigh, "Scientific Papers," vol. I, 1899. "Investigations in optics, with special reference to the spectroscope," pp. 415–459, (1879–1880). (The resolving power of an aperture, p. 420.)

(2) R. V. L. Hartley, "Relations of carrier and sidebands in radio transmission," Proc. I.R.E., vol. 11, pp. 34–56; February, (1923). (The frequency spectrum of a transient signal.)

(3) H. Nyquist, "Certain factors affecting telegraph speed," *Bell Sys. Tech. Jour.*, vol. 3, pp. 324–346; April, (1924). (The frequency spectrum of a transient signal, and a shaping filter.)

(4) H. E. Ives, "Television," *Bell Sys. Tech. Jour.*, vol. 6, pp. 551–559; October, (1927). (The required width of frequency band.)

(5) F. Gray, J. W. Horton, and R. C. Mathes, "The production and utilization of television signals," *Bell Sys. Tech. Jour.*, vol. 6, pp. 560–603; October, (1927). (Aperture distortion and frequency spectrum, the equivalent filter characteristics, pp. 573–582, 600–603.) (In Fig. 17, curves 1 and 2 should have their numbers interchanged.)

(6) H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. A.I.E.E.*, vol. 47, pp. 617–644; April, (1928). (Relations between filter characteristics and the shape of a transient signal.)

(7) R. V. L. Hartley, "Transmission of information," *Bell Sys. Tech. Jour.*, vol. 7, pp. 535–563; July, (1928). (The basic theory of frequency band-width requirements.)

(8) G. A. Campbell and R. M. Foster, "Fourier integrals for practical applications," Bell Telephone System Technical Publications, Monograph B-584, September, (1931). Shorter edition, *Bell Sys. Tech. Jour.*, vol. 7, pp. 639–707; October, (1928). (Introduction to the Fourier integral, and extensive tables.)

(9) J. C. Wilson, "The design of television transmission equipment," *Jour. Television Soc.*, series 2, vol. 1, pp. 237–271; September, (1933). (Determination of the frequency-band requirements, pp. 247–253.)

(10) E. W. Engstrom, "A study of television image characteristics," Proc. I.R.E., vol. 21, pp. 1631–1651; December, (1933). (The picture definition required by the eye.)

(11) V. K. Zworykin, "Description of an experimental television system and the kinescope," Proc. I.R.E., vol. 21, pp. 1655–1673; December, (1933). (The intensity distribution of a cathode-ray spot, Fig. 7.)

(12) P. Mertz and F. Gray, "A theory of scanning and its relation to the characteristics of the transmitted signal in telephotography and television," *Bell Sys. Tech. Jour.*, vol. 13, pp. 464–515; July, (1934). (Two-dimensional analysis of aperture distortion and the frequency spectrum.)

(13) R. D. Kell, A. V. Bedford, and M. A. Trainer, "An experimental television system—the transmitter," Proc. I.R.E., vol. 22, pp. 1246–1265; November, (1934). (Formulas for width of frequency band and optimum number of scanning lines, pp. 260–261; the correction factor for the resultant width of confusion of the scanning lines in transmitter and receiver.)

(14) I. G. Maloff and D. W. Epstein, "Theory of electron gun," Proc. I.R.E., vol. 22, pp. 1386–1411; December, (1934). (The intensity distribution of a cathode-ray spot; Maxwellian distribution, pp. 318–319; electron lens aberration, pp. 329–330.)

(15) A. T. Starr, "Electric Circuits and Wave Filters," 1934. Chap. XII, "Transients in networks," pp. 332–353. (Transient distortion in an idealized low-pass filter.)

(16) E. A. Guillemin, "Communication Networks," vol. II, 1935. Chap.

XI, "The transient behavior of filters," pp. 461–507. (Transient distortion in idealized low-pass and band-pass filters.)

(17) D. W. Epstein, "Electron optical system of two cylinders as applied to cathode-ray tubes," PROC. I.R.E., vol. 24, pp. 1095–1139; August, (1936). (Electron lens aberration, pp. 444–452.)

(18) J. C. Wilson, "Television Engineering," 1937. Chap. IV, "Analysis of finite aperture scanning methods," pp. 73–108. Chap. XII, "Physical limitations," pp. 420–437. (Many references listed.) (Aperture distortion and frequency characteristics.) (Frequency band-width requirements.)

(19) L. C. Jesty and G. T. Winch, "Television images, an analysis of their essential qualities," *Trans. Ill. Eng. Soc.* (London), Communication No. 217, Aug., 1937. (Transverse intensity distribution of cathode-ray scanning lines.)

(20) A. V. Bedford, "Figure of merit for television performance," *R.M.A. Eng.*, vol. 2, pp. 5–7; November, (1937). (The observation of average horizontal and vertical resolution in terms of the number of lines in a distance equal to the height of the image.)

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

CONCENTRIC NARROW-BAND-ELIMINATION FILTER*

By

Laurance M. Leeds

(Radio Transmitter Engineering Department, General Electric Company, Schenectady, New York)

Summary—*A type of wave filter employing a concentric transmission line is described. Filters of this type are especially applicable to ultra-high-frequency operation where their physical size becomes reasonable and where the conventional lumped-constant filters become practically unrealizable.*

It is shown that the addition of a simple lumped-reactance element will result in a sharp cutoff characteristic analogous to that exhibited by a conventional m-derived lumped-constant filter but over a limited frequency band. This latter filter has allowed simultaneous transmission and reception using the same antenna, in the 30- to 42-megacycle band, transmitter power 150 watts, receiver sensitivity 5 microvolts, with a frequency difference of only 0.9 megacycle.

It is shown that a 14 per cent increase in driving-point impedance of the filter can be obtained if the filter line is shortened 0.065 wavelength and tuned to resonance with a shunt capacitance. This result is applicable directly to the use of transmission lines as tank circuits when very high driving point impedance is required.

PROGRESS in the electrical-communications art depends to a great extent upon the ability to obtain selective frequency discrimination. Although the theory and design procedure of lumped-constant wave filters is well established, the physical realization of the desired characteristic becomes more difficult as the frequency is increased. This difficulty is due, primarily, to the high coil dissipation and low Q which inhibit the attainment of sharp cutoff characteristics and reasonable attenuation. The high distributed capacitance of the coils and stray capacitances of the circuit intensify the difficulties encountered.

Much progress has been made in the application of quartz crystals to wave filters, operating at very low potentials, to attain high attenuation and sharp cutoff characteristics. However, this type of filter is not applicable to ultra-high frequencies because of crystal frequency limitations.

Another type of filter construction by which the required reactive elements may be realized at the ultra-high frequencies should be noted in passing. This involves the use of sections of concentric transmission line as the reactive elements. The principle utilized here is simply that a section of line of a length other than its resonant-frequency length exhibits a reactive component.

* Decimal classification: R386. Original manuscript received by the Institute, October 25, 1937; revised manuscript received by the Institute, January 19, 1938. Presented before Silver Anniversary Convention, May 12, 1937, New York City.

The present paper describes a narrow-band or peak-elimination filter which is especially applicable to ultra-high-frequency operation. Although a concentric transmission line is employed, a mode of operation is achieved which must be distinguished from that previously alluded to. In this filter the phenomenon of standing waves on a transmission line is taken advantage of to obtain the desired elimination characteristic. An element of simplicity enters into the filter in that a single section of transmission line is used thereby permitting initial tuning and adjustment to be accomplished simply by a variation of the physical length.

The performance of these filters is closely analogous, over a *limited* frequency range, to that obtained from the customary filter theory of Campbell, Shea, and others, just the same as the impedance characteristics of an antenna may be represented by a lumped-constant circuit over a limited frequency range. The multiple-resonance effects of antennas and transmission lines prevent the analogy over an extended frequency range. The important difference between the two filter systems is that the concentric filter can easily be realized physically at the ultra-high frequencies and will yield very high attenuations, whereas the realization of the lumped-constant filter is extremely difficult and the performance of the best attainable is far from satisfactory.

The concentric filter comprises a section of transmission line of a particular length with output terminals at a suitable distance from the remote end, which may either be open- or short-circuited, dependent upon design. The potential distribution for the half-wave filter is shown by (a) of Fig. 1. The remote end, that on the right, is open-circuited. The standing voltage wave at the resonant frequency, which is also the frequency of maximum attenuation, is indicated by F_T. If the output of the filter is terminated in an impedance Z_o equal to the surge impedance of the filter line, then the standing voltage waves at frequencies other than at resonance may be indicated by F_L, which is less than F_T, and F_H which is greater than F_T. In the first-quarter wave these are shown as practically constant since, at frequencies off resonance, the line between input and output is substantially terminated in its surge impedance. At the elimination frequency F_T, the impedance of the line at the output terminals is extremely low. This results in the output potential, at the elimination frequency undergoing attenuation of the order of 50 decibels, dependent upon filter design. This will be shown later in the performance characteristics.

A filter one wavelength long at the resonant frequency and open-circuited at the remote end is shown by (b) of Fig. 1. As in the half-

wavelength filter the potentials at the remote end are a maximum since the line is discontinuous at this point, and the potential distribution arguments are the same as those already given. The essential operational difference between the two filters is that the shorter exhibits somewhat greater peak and wider attenuation-versus-frequency characteristic.

If the filter is short-circuited at the remote end, the shortest length is three quarters of a wavelength as shown by (c) of Fig. 1. In this case

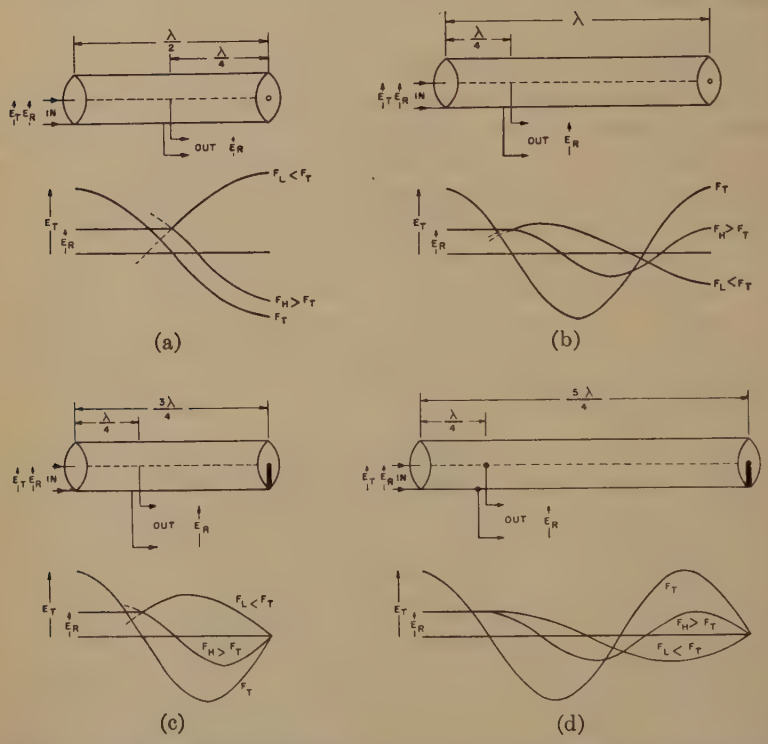


Fig. 1—Potential distributions for various length concentric filters, receiver impedance equal to z_0.
(a) λ/z long, open-circuited at remote end.
(b) λ long, open-circuited at remote end.
(c) $3\lambda/4$ long, short-circuited at remote end.
(d) $5\lambda/4$ long, short-circuited at remote end.

the potentials at the remote end, or reflection point, must be zero, and the remaining distribution as previously discussed.

The next longer resonant length for a line short-circuited at the remote end is five fourths of a wavelength as shown by (d) of Fig. 1. Observe that the output terminals could be made at a point one-half

wavelength from the remote end where the potential at the elimination frequency F_T is practically zero. If the output terminals at this latter point are terminated in a receiving impedance equal to the surge impedance Z_o of the line the resulting characteristic would have the lower peak attenuation of the five-fourths-wavelength filter with the broader attenuation band of the three-quarters-wavelength filter.

Before investigating the characteristics of the filter consider a practical application for which it was initially developed. To provide for duplex operation it has been customary to use separate transmitting and receiving antennas supported on individual towers or separated by an outrigger on one tower. However, a suitable filter to keep the transmitter energy out of the receiver would permit simultaneous transmission and reception using the same antenna, materially

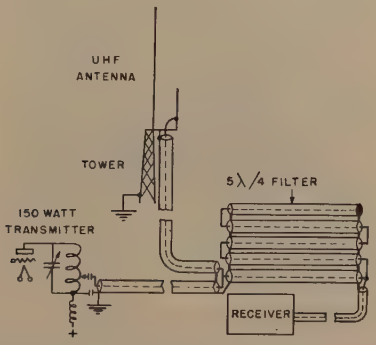


Fig. 2—Block diagram of typical police-headquarters installation using concentric filter for duplex operation on a single antenna.

simplifying the structures required. A typical police-headquarters installation operating in the 30- to 42-megacycle band and using a concentric filter for duplex operation on the same antenna is shown in Fig. 2. Incidentally, a filter is also used in the mobile equipment to provide duplex operation using the single rod antenna. Fig. 3 shows the one-wavelength filter used in conjunction with a 15-watt mobile transmitter. A trombone section on the remote end is used for initial tuning and the receiver tap-off is made through flexible concentric line. Prior to the introduction of turret-top cars the customary procedure was to use the roof antenna for reception and a vertical rod for transmission. Tests have shown, however, that use of the filter and vertical-rod antenna for reception has resulted in a signal-strength increase of from two to five times due, presumably, to the greater effective height of the rod antenna and the vertical polarization of the headquarters transmitted wave.

Since the transmitter and receiver operate on fixed frequencies the filter must possess the following characteristics:

1. It must present a high input impedance at transmitter frequency so as not to load the transmitter. Any power taken by the filter is dissipated as heat which causes physical elongation resulting in a tuning shift with loss of attenuation.
2. It must have sufficient attenuation at transmitter frequency so as not to overload the first tube of the receiver. The normal receiver selectivity will then be sufficient to reject the signal from the second detector.
3. At the receiver frequency it must have low attenuation and present an input impedance comparable to that of the antenna transmission line.

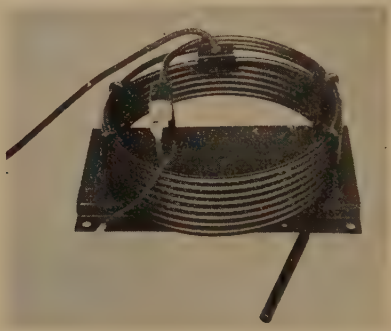


Fig. 3—One-wavelength concentric filter employing a 3/8-inch transmission line and used in conjunction with a 15-watt ultra-high-frequency transmitter for duplex operation on a single antenna.

The frequency-attenuation characteristic of a five-fourths-wavelength filter using three-quarter-inch copper tubing is shown in Fig. 4. At the resonant frequency of 35.6 megacycles, which is approximately in the middle of the ultra-high-frequency police band, the filter exhibits an attenuation of 41 decibels.

For fixed values of filter length, outer-tube diameter, and frequency, the input impedance is dependent upon the diameter ratio b/a of the line conductors, as shown by Fig. 5. It is a maximum at 9.2 which corresponds to a surge impedance of 133 ohms. The peak attenuation, which is dependent upon a different function of the line attenuation constants, is maximum at a diameter ratio of 3.6. Since in the particular application considered, a considerable amount of power is available at the elimination frequency, the diameter ratio for maximum input impedance was chosen. Because of conveniently available tube and wire sizes the actual diameter ratio employed was 8.57 instead of 9.2. The difference in input impedance is negligible. If the tube diam-

eters are varied simultaneously, maintaining a constant diameter ratio, other factors remaining unchanged, the decibel attenuation

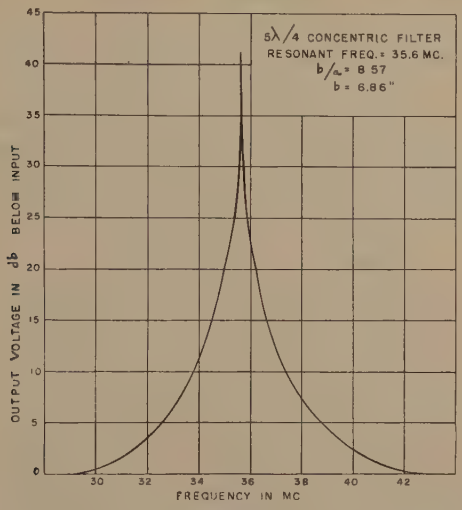


Fig. 4—Attenuation versus frequency for fixed geometry, receiver impedance equal to z_0.

varies almost directly as the logarithm of the inner diameter of the outer tube as shown by Fig. 6. Obviously the selection of tube size is also governed by certain economies and physical size limitations.

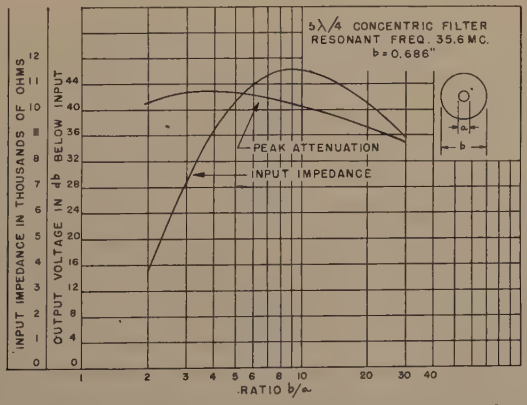


Fig. 5—Peak attenuation and input impedance for various diameter ratios.

Because of the low attenuation constant of a well-constructed concentric line the change in input resistance at frequencies close to resonance is very rapid. This is shown by Fig. 7, wherein the in-

put resistance of a five-fourths-wavelength filter is plotted against frequency. The ordinate is resistance-per-unit of surge impedance. Since the line surge impedance is 129 ohms for a diameter ratio of 8.57, the peak value of 90.1 corresponds to 11,620 ohms. The upper curve, showing the attenuation when the filter output is terminated in an impedance equal to the surge impedance, is the normal mode of operation. It may be noted that the input resistance at pass or receiver frequencies is approximately 0.7 Z_o or about 90 ohms. Fig. 8 shows the resistive and reactive components of the input impedance for the same filter. The rapid variation of resistance and reactance in

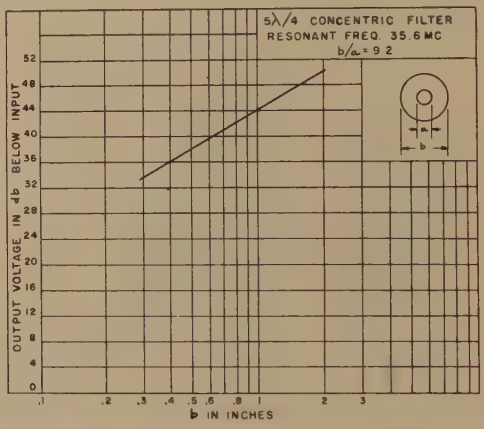


Fig. 6—Peak attenuation versus variation of outer-tube diameter for optimum b/a ratio for maximum input impedance.

the region of the resonant frequency is clearly indicated. It should be observed that some input reactance remains even at pass frequencies 10 per cent removed from the elimination frequency.

If the input reactance at any particular pass frequency is tuned out by means of a series element an attenuation characteristic results which is analogous, over a limited frequency range, to that of the m-derived lumped-constant filter. Fig. 9 shows the attenuation characteristic of a half-wavelength filter resonant at 35.6 megacycles. The full-line curve gives the attenuation for the case where the input reactance is zero at 34.8 megacycles. For any other receiver frequency at which the input reactance is made zero the minimum attenuation is that indicated by the dotted locus curve. The greater peak attenuation of 53 decibels is due to the filter length being only a half wavelength. It is interesting to note that, in general, the attenuation at a particular pass frequency is less than zero decibels thereby constituting a transmission gain. This is a consequence of the driving-point imped-

ance being lower than the output-load impedance. For receiver frequencies less than the elimination frequency the input reactance of the filter proper is inductive necessitating the use of a series capacitor. A similar attenuation curve for a pass frequency greater than the

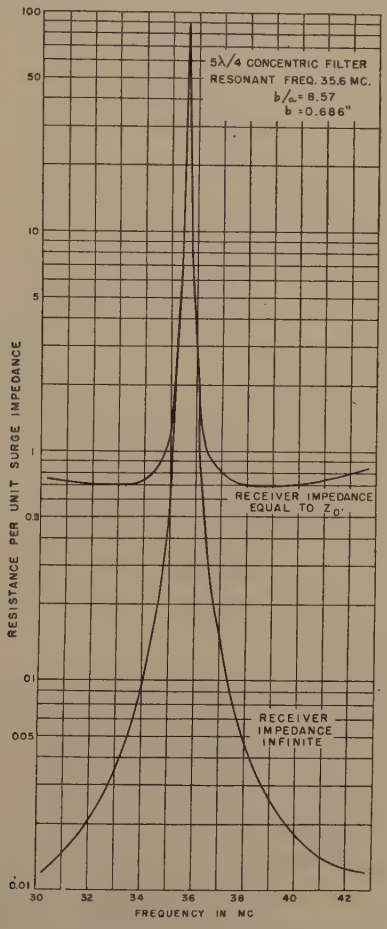


Fig. 7—Input resistance of 5λ/4 concentric filter.

elimination frequency will, of course, be obtained, but a series inductance is required to obtain zero-input reactance. The peak attenuation at the elimination frequency is unaffected by the various series reactances since the resistive component of the input impedance is approximately 29,000 ohms—very much greater than any reactance which may be called for.

It should be observed that if a transmission line of a length suit-

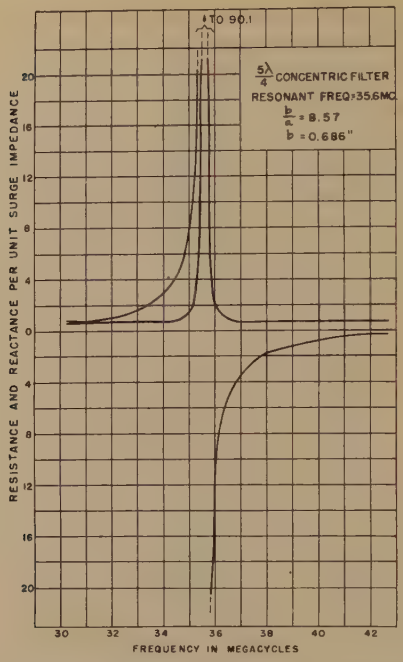


Fig. 8.—Input impedance of $5\lambda/4$ concentric filter, receiver impedance equal to z_0.

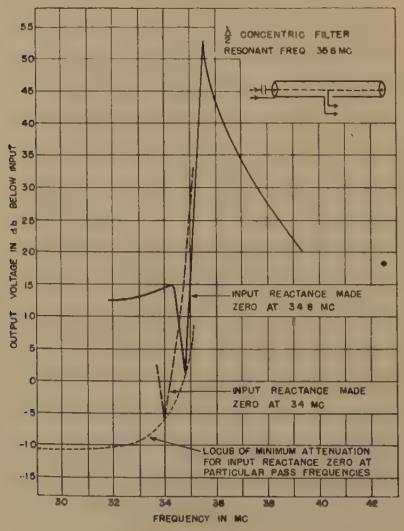


Fig. 9—Attenuation characteristic of $\lambda/2$ concentric filter for pass frequencies below resonance—input reactance zero at any particular pass frequency.

able to exhibit antiresonance is physically shortened and retuned by means of a shunt capacitor, the input impedance may increase for certain shortening lengths. Fig. 10 shows graphically the improvement in driving-point impedance that will be obtained for various length lines. For example, a quarter-wavelength line ($n=1$) short-circuited at the remote end will exhibit 14 per cent more input resistance if the line is physically shortened 0.065 wavelength and tuned to antiresonance by means of a shunt capacitor. This method of line tuning has advantages over trombone sections for certain applications. By a judicious choice of line length it is possible to design a half-wavelength

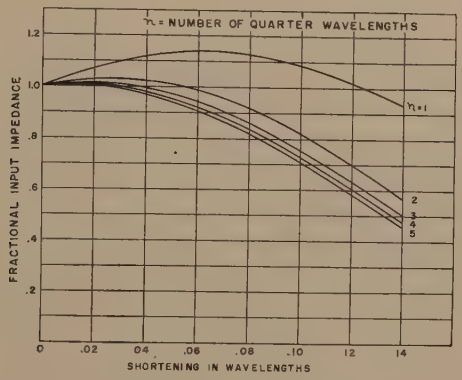


Fig. 10—Input impedance of shortened line tuned by shunt capacitance.

filter which exhibits the sharp cutoff characteristic of the *m*-derived filter and which may be continuously adjusted to any elimination and pass frequency in the 30- to 42-megacycle band.

To summarize briefly, concentric filters may be constructed for operation at the ultra-high frequencies where lumped-constant filters are practically unrealizable. Peak attenuations of considerable magnitude may be obtained from a single filter and, of course, more than one filter may be used in combination. They have demonstrated their utility in police-communication systems by allowing simultaneous transmission and reception on the same antenna.

Appendix

Terminology

Subscript s refers to sending ends
Subscript r refers to receiving ends
Z_0 = line surge impedance
Z_r = receiver impedance

γ = complex propagation constant
p = ratio of Z_r to Z_o
L = total line length in wavelengths at resonant frequency
L_1 = line length in wavelengths from receiver tap to remote end at resonant frequency
L_2 = line length in wavelengths from sending end to receiver tap at resonant frequency
F = frequency in megacycles
F_T = resonant frequency in megacycles
$\delta = F/F_T$ = fractional resonant frequency

A. Evaluation of Equations

In order to condense the Appendix many of the equations given below are in terms of the complex propagation constant γ. The evaluation of these equations may be accomplished by first expanding the functions of the complex argument in terms of their real and imaginary components. Since $\gamma L = \alpha L + j\beta L = (R/2z_0)L + j(2\pi/\lambda)L$ the resulting functions will have real arguments.

The evaluation of the attenuation constant for a concentric line may be carried out as follows:

Using A. Russell's asymptotic formula[1] for the high-frequency resistance the attenuation constant expressed in practical units becomes

$$\alpha = \frac{R}{2Z_0} = \frac{0.5\sqrt{F}\left(\frac{b}{a}+1\right)\times 10^{-3}}{b\ 138\ \log_{10}\left(\frac{b}{a}\right)} \text{ nepers/foot} \qquad (1)$$

a = outer diameter of inner conductor in inches
b = inner diameter of outer conductor in inches

$$\alpha L = \frac{0.3566\times 10^{-3}\left(\frac{b}{a}+1\right)\sqrt{\delta}}{\sqrt{F_T}\, b \log_{10}\left(\frac{b}{a}\right)} \cdot \frac{L}{\lambda} \text{ nepers.} \qquad (2)$$

B. Filters Odd Number of Quarter Wavelengths Long Short-Circuited at Remote End

Ratio of receiving- to sending-end voltage for fractional resonant frequencies

$$\frac{E_r}{E_s} = \frac{p \sinh(\gamma L_1 \delta)}{p \sinh(\gamma L \delta) + \sinh(\gamma L_1 \delta) \sinh(\gamma L_2 \delta)} . \qquad (3)$$

[1] Sterba and Feldman, "Transmission lines for short-wave radio systems," PROC. I.R.E., vol. 20, p. 1167; July, (1932).

Filter input impedance

$$Z_s = Z_0\left[\frac{p \sinh(\gamma L\delta) + \sinh(\gamma L_1\delta)\sinh(\gamma L_2\delta)}{p\cosh(\gamma L\delta) + \sinh(\gamma L_1\delta)\cosh(\gamma L_2\delta)}\right]. \quad (4)$$

C. Filters Even Number of Quarter Wavelengths Long Open-Circuited at Remote End

Ratio of receiving- to sending-end voltage for fractional resonant frequencies

$$\frac{E_r}{E_s} = \frac{p\cosh(\gamma L_1\delta)}{p\cosh(\gamma L\delta) + \cosh(\gamma L_1\delta)\sinh(\gamma L_2\delta)}. \quad (5)$$

Filter input impedance

$$Z_s = Z_0\left[\frac{p\cosh(\gamma L\delta) + \cosh(\gamma L_1\delta)\sinh(\gamma L_2\delta)}{p\sinh(\gamma L\delta) + \cosh(\gamma L_1\delta)\cosh(\gamma L_2\delta)}\right]. \quad (6)$$

D. Half-Wavelength Filter Open-Circuited at Remote End—Receiver Impedance Equal to Line Surge Impedance—Driving-Point Reactance Tuned out by Series Element at Receiver Frequency

Input impedance for fractional resonant frequencies

The driving-point impedance of the filter line without a series element is given by (7).

$$Z_s' = Z_0\left[\frac{2\cosh(\gamma L\delta) + \sinh(\gamma L\delta)}{2\sinh(\gamma L\delta) + \cosh(\gamma L\delta) + 1}\right]. \quad (7)$$

Denote the real part of Z_s' by AZ_0 and the reactive part by $-jkZ_0$. (The use of the negative sign results in the sign of k itself determining whether a series inductance or capacitance is required.) By suitable expansion in terms of functions of the complex argument the following equations result:

$$A = \frac{1}{D}\Big(2\cosh(\alpha L\delta) + \sinh(\alpha L\delta)\Big)\Big(\cos(\beta L\delta) + 2\sinh(\alpha L\delta) + \cosh(\alpha L\delta)\Big) \quad (8)$$

$$k = \frac{1}{D}\Big(3\sin(\beta L\delta)\cos(\beta L\delta) - \sin(\beta L\delta)\left[2\sinh(\alpha L\delta) + \cosh(\alpha L\delta)\right]\Big) \quad (9)$$

where D is given by (10)

$$D = \left[\begin{array}{l} 3\sin^2(\beta L\delta) + 3\sinh^2(\alpha L\delta) + \cosh(\alpha 2L\delta) \\ + 2\sinh(\alpha 2L\delta) \\ \quad + 2\cos(\beta L\delta)\left[2\sinh(\alpha L\delta) + \cosh(\alpha L\delta)\right] + 1 \end{array}\right]. \quad (10)$$

The solution of (9) for $\delta = \delta_r = F_r/F_T$ gives the value of reactance required to make the input impedance resistive at receiver or pass frequency F_r.

Let k_r denote the value of (9) for $\delta = \delta_r$. The input impedance may then be expressed by (11) or (12) depending upon the ratio F_r/F_T.

$$Z_s = Z_0[A - j(k - k_r\delta_r/\delta)] \qquad F_r \leqq F_T \tag{11}$$

$$Z_s = Z_0[A - j(k - k_r\delta/\delta_r)] \qquad F_r \geqq F_T. \tag{12}$$

Ratio of receiving- to sending-end voltage for fractional resonant frequencies

$$\frac{E_r}{E_s} = \cosh(\gamma L_2\delta) - (1/Z_s)\sinh(\gamma L_2\delta) \tag{13}$$

wherein Z_s is given by (11) or (12), whichever applies.

E. Resonant Input Impedance of a Physically Shortened Line Tuned to Resonance at the Input End with a Shunt Capacitance[2]

The driving-point impedance of a line which has been physically shortened A wavelengths from an odd number of quarter wavelengths long and short-circuited at the remote end or an even number of quarter wavelengths long and open-circuited at the remote end, and which has lumped shunt capacitance at the input terminal is given by the following equation:

$$Z_s = Z_0\left[\frac{\tanh(\alpha L)\sec^2\left(\frac{2\pi A}{\lambda}\right) + j\left[\tan\left(\frac{2\pi A}{\lambda}\right) - \frac{Z_0}{X_c} - \tanh^2(\alpha L)\tan\left(\frac{2\pi A}{\lambda}\right)\left(1 + \frac{Z_0}{X_c}\tan\left(\frac{2\pi A}{\lambda}\right)\right)\right]}{\left(\tan\left(\frac{2\pi A}{\lambda}\right) - \frac{Z_0}{X_c}\right)^2 + \tanh^2(\alpha L)\left(1 + \frac{Z_0}{X_c}\tan\left(\frac{2\pi A}{\lambda}\right)\right)^2}\right]. \tag{14}$$

For the shunt capacitor to resonate the line the reactive part is set equal to zero.

$$X_c = Z_0\left[\frac{1 + \tan^2\left(\frac{2\pi A}{\lambda}\right)\tanh^2(\alpha L)}{\tan\left(\frac{2\pi A}{\lambda}\right)(1 - \tanh^2(\alpha L))}\right]. \tag{15}$$

[2] In the interim between presentation of this paper and submission for publication the increase in driving-point impedance which can be obtained by judiciously shortening the line and tuning to resonance by a shunt capacitor has been shown by another investigator, Salzberg, "On the optimum length for transmission lines used as circuit elements," Proc. I.R.E., vol. 25, pp. 1561–1564; December, (1937).

For well-constructed radio-frequency lines $\tanh(\alpha L) \ll 1$. Likewise for A less than approximately $\lambda/6$ the bracketed term of the numerator can be set equal to one. With these simplifications

$$X_c \cong Z_0 \cot\left(\frac{2\pi A}{\lambda}\right). \tag{16}$$

The fractional input impedance of the resonated shortened line to that of the unshortened line can be expressed, using the same simplification, as follows:

$$\frac{Z_{s1}}{Z_{s0}} = \frac{n \cos^2\left(\frac{2\pi A}{\lambda}\right)}{n - \frac{4A}{\lambda}} \tag{17}$$

where n is the length of the unshortened line in quarter wavelengths.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

THEORY OF THE DISCRIMINATOR CIRCUIT FOR AUTOMATIC FREQUENCY CONTROL*

By

Hans Roder

(General Engineering Laboratory, General Electric Company, Schenectady, New York)

Summary—*The theory of the frequency-discriminator circuit which is used in radio receivers equipped with automatic frequency control is worked out. Formulas for computing gain, frequency-control sensitivity, adjacent-channel attentuation, pull-in range, etc., are developed. These formulas apply to any value of primary or secondary inductance and to any value of coil quality factor. The loading represented by the plate impedance of the driver tube and by the diodes is also taken into account. After slight modification, the resulting formulas can be used for various different circuit arrangements. Nonlinear audio-frequency distortion is found to exist in the case where the channels for signal and automatic frequency control are common; these distortions, however, are small.*

It is found that the loading effect of the diodes is responsible for the poor selectivity resulting for most automatic-frequency-control transformers. A circuit in which the automatic frequency control and the signal channel are separated is preferable if better selectivity is desired.

DURING 1936 a number of radio-set manufacturers brought out radio receivers with automatic frequency control utilizing a principle developed by the RCA License Laboratories.[1,2] These receivers incorporate an interesting circuit which was named "Phasen—Sprung—Schaltung" by its inventor, Hans Riegger, in 1922. Since that time, Riegger's circuit is being used for automatic frequency control of the high-frequency alternators at the Nauen Station.[3,4,5,6,7]

Foster and Seeley have recently published a detailed description of the automatic-frequency-control (AFC) principle as applied to radio receivers,[2] referring to the theory and operation of the discriminator circuit and of the automatic-frequency-control tube. In their theory of the discriminator circuit some simplifying assumptions have been made.

* Decimal classification: R360. Original manuscript received by the Institute, April 16, 1937; revised manuscript received by the Institute, February 3, 1938.

[1] Charles Travis, "Automatic frequency control," Proc. I.R.E., vol. 23, pp. 1125–1141; October, (1935).

[2] D. E. Foster and S. W. Seeley, "Automatic tuning, simplified circuits, and design practice," Proc. I.R.E., vol. 25, pp. 289–313; March, (1937).

[3] H. Riegger, German Patent, 358,877; German Patent, 413,622; German Patent, 428,643.

[4] H. Riegger, Scientific Publication, Siemens Concern, vol. 1, p. 126, (1920).

[5] P. Schuchmann, German Patent, 490,588.

[6] P. Schuchmann, *Telefunken-Zeit.*, no. 40–41, p. 29.

[7] H. Roder, "Handbook of Electricity and Magnetism," edited by L. Graetz, Munich, 1928, p. 781.

In the following paper a complete theory of the automatic-frequency-control discriminator circuit is presented, resulting in formulas which may be used to calculate gain, automatic-frequency-control sensitivity, pull-in width, adjacent-channel attenuation (ACA), etc. These formulas bear a very close relationship to those for a standard two-tuned-circuit intermediate-frequency transformer. It therefore appears advisable first to give a short account of the theory of the standard intermediate-frequency transformer.

I. The Standard Intermediate-Frequency Transformer

We replace the driver tube with its "constant-current—parallel-R_p" network and thus obtain the circuit of Fig. 1. Both primary and sec-

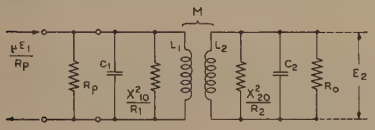


Fig. 1—Illustration of nomenclature.

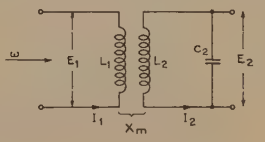


Fig. 2

ondary circuit shall be tuned to the carrier frequency ω_0, so that

$$\omega_0 L_1 = \frac{1}{\omega C_1} = X_{10}$$

$$\omega_0 L_2 = \frac{1}{\omega_0 C_2} = X_{20}.$$

The parallel resistors $X_{10}^2/R_1 = X_{10}Q_1$ and $X_{20}^2/R_2 = X_{20}Q_2$ account for any resistances in the tuned circuits (coil resistance, plus losses caused by the shield can, plus losses in the tuning capacitors). The resistor R_0 is infinite if the secondary is working into the grid of the next tube; it is approximately $\frac{1}{2}R_{dc}$, if it works into a diode whose direct-current load resistance is R_{dc}. Since we consider only a narrow band of frequencies we may put

$$\frac{\omega}{\omega_0} = 1 + \delta$$

where,

$$\delta = \frac{\omega - \omega_0}{\omega_0} = \frac{\text{modulation (audio) frequency}}{\text{carrier frequency}}. \qquad (1)$$

Then

$$\omega L - \frac{1}{\omega C} = X_0\left(1 + \delta - \frac{1}{1+\delta}\right) = X_0 2\delta. \qquad (2)$$

The reactances of L_1, L_2, M, C_1, and C_2 can be assumed to be constant over the narrow frequency range considered. With these simplifications we find for the *gain* of the stage

$$\text{gain} = \frac{E_2}{E_1} = +j\frac{\mu}{R_p}\,\frac{X_{m0}}{(p_1 + j2\delta)(p_2 + j2\delta) + k^2} \tag{3}$$

$$= +j\frac{\mu}{R_p}\,\frac{X_{m0}}{p_1p_2 + k^2} \text{ at resonance } (\delta = 0)$$

where,

$$X_{m0} = \omega_0 M$$

$$\frac{X_{m0}^2}{X_{10}X_{20}} = k^2 = \text{coupling coefficient}$$

$$p_1 = \frac{1}{Q_1} + \frac{X_{10}}{R_p} = \text{primary power factor} \tag{4}$$

$$p_2 = \frac{1}{Q_2} + \frac{X_{20}}{R_0} = \text{secondary power factor.} \tag{5}$$

The *attenuation* A provided by the circuit for a frequency deviation of δ is given by

$$A = \frac{\text{gain at } \delta = 0 \text{ (resonance)}}{\text{gain at frequency deviation of } \delta}$$

$$= \frac{\sqrt{(p_1p_2 + k^2 - (2\delta)^2)^2 + (2\delta(p_1 + p_2))^2}}{p_1p_2 + k^2}. \tag{6}$$

Band width is defined as the frequency deviation $(2\delta)_b$ at which $A^2 = 2$; i.e., at which E_2 is equal to 0.707 of its value at resonance.

The adjacent-channel attenuation at 10-kilocycle frequency deviation is found by computing A for

$$(2\delta)_a = \frac{10 \text{ kilocycles}}{\text{intermediate frequency}}.$$

Critical coupling[8] is defined as the tightest coupling which still will result in a *single-peak* resonance curve for E_2. It is found to be

$$k^2_{\text{crit}} = \tfrac{1}{2}(p_1^2 + p_2^2). \tag{7}$$

[8] This definition of "critical" coupling is different from that used by G. W. Pierce in his book "Electric Waves and Oscillations," published by the McGraw-Hill Book Company, New York, 1920, p. 156. Pierce's "critical" coupling is $X_m^2 = R_1, R_2$ or, in our terminology, $k^2 = 1/Q_1Q_2$. It is the coupling for maximum energy transfer.

At critical coupling

$$\text{resonance gain} = +j\frac{\mu}{R_p}\,\frac{\sqrt{2(p_1^2+p_2^2)X_{10}X_{20}}}{(p_1+p_2)^2} \tag{8}$$

$$\text{band width} = (2\delta)_b = \frac{1}{\sqrt{2}}(p_1+p_2) \tag{9}$$

$$\text{adjacent-channel attenuation} = \sqrt{1+4\left(\frac{2\delta_a}{p_1+p_2}\right)^4}\,. \tag{10}$$

The expressions for the power factors (equations (4) and (5)) make allowance for the loading of the primary by the driver-tube plate resistance and for the loading of the secondary due to the diode.

Formulas (3) to (10) contain all information required for the design of intermediate-frequency transformers.

II. Theory of the Automatic-Frequency-Control Discriminator Circuit

A. Principle

The currents in two coupled circuits, in which the secondary circuit is tuned to the operating frequency, are 90 degrees out of phase as seen from the following consideration. Referring to Fig. 2, we have

$$E_{12} = jX_mI_1 \text{ and}$$

$$I_2 = \frac{E_{12}}{R_2+j(X_{L2}-X_{c2})} = \frac{jX_LI_1}{R_2+j(X_{L2}-X_{c2})}\,.$$

At secondary resonance, $X_{L2}=X_{c2}$, hence I_2 becomes 90 degrees leading (or lagging, dependent on polarity of X_m) with respect to I_1. It does not matter whether or not the primary is tuned. Correspondingly, the voltages E_1 and E_2 have a 90-degree phase relation.

Next, consider the circuit of Fig. 3. The primary and secondary circuits, together with two rectifiers and a load resistor, form a bridge circuit. The radio-frequency voltage for the No. 1 rectifier is $E_1-\frac{1}{2}E_2$; that for the No. 2 rectifier is $E_1+\frac{1}{2}E_2$. The direct-current potential developed on the load resistor between BC is proportional to the vector sum $E_1+\frac{1}{2}E_2$; the potential between AC is proportional to $E_1-\frac{1}{2}E_2$.

Fig. 4 shows the vector diagram for the voltages $E_1\pm\frac{1}{2}E_2$, for a constant voltage E_1, but variable frequency ω. If $\omega=\omega_{20}$ (which latter is the resonant frequency of the secondary), then $E_1+\frac{1}{2}E_2$ and $E_1-\frac{1}{2}E_2$ are equal, because of the 90-degree phase relation between E_1 and E_2.

Consequently, the potential difference between AB is zero. If ω deviates from ω_{20}, the voltages $E_1+\frac{1}{2}E_2$ and $E_1-\frac{1}{2}E_2$ will be different, resulting in a potential difference between A and B. Both the polarity and the magnitude of the voltage AB depend, within limits, upon the sign and magnitude of the frequency deviation from the nominal value ω_{20}. The voltage AB may consequently be used to operate a "frequency control" tube in order to provide automatic tuning.

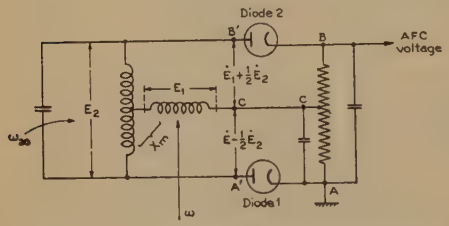


Fig. 3—Fundamental automatic-frequency-control discriminator circuit.

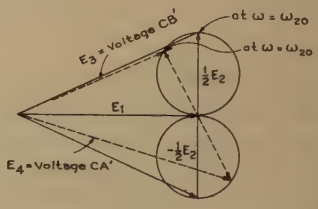


Fig. 4

B. Analysis of the Circuit

For the analysis, we shall use the circuit shown in Fig. 5. Since this circuit is a modification of that considered in Fig. 3 it is most convenient for analytical work. It is not the circuit which has been used in practice, but it will be seen that the resulting formulas can be easily modified, to include that circuit and others.

Here the direct currents of the two rectifiers do not interact with each other but each rectifier works into a resistance of $\frac{1}{2}R_{dc}$, where R_{dc} is the total resistance between A and B. The equivalent radio-frequency load which each rectifier and load resistor reflects upon the circuit is equal to that represented by a resistance

$$R_0 = \tfrac{1}{2}(\tfrac{1}{2}R_{dc}).$$

The nomenclature is illustrated in Fig. 5. Both the primary and the secondary circuit are assumed to be tuned to the frequency ω_0. The primary circuit impedance is (equation (2)),

$$X_{10}\left(\frac{1}{Q_1}+j2\delta\right),$$

and the secondary impedance is

$$X_{20}\left(\frac{1}{Q_2}+j2\delta\right).$$

α_1 and α_2 are the transformer ratios. The following approximations can be made:

$$jX_{L_1} \ll R_p,$$
$$(-jX_{10} + \tfrac{1}{2}jX_{L2}) \ll R_0.$$

Using these approximations, we write down the mesh equations as follows:

I_0	I_1	I_2	I_3	I_4
R_p	$+j\alpha_1 X_{10}$	0	0	$0=\mu E_0$
$+j\alpha_1 X_{10}$	$X_{10}\left(\frac{1}{Q_1}+j2\delta\right)$	$+jX_m$	$+jX_{10}$	$-jX_{10}=0$
0	$+jX_m$	$X_{20}\left(\frac{1}{Q_2}+j2\delta\right)$	$-j\frac{1}{2}\alpha_2 X_{20}$	$-j\frac{1}{2}\alpha_2 X_{20}=0$
0	$+jX_{10}$	$-j\frac{1}{2}\alpha_2 X_{20}$	R_0	$0=0$
0	$-jX_{10}$	$-j\frac{1}{2}\alpha_2 X_{20}$	0	$R_0=0$

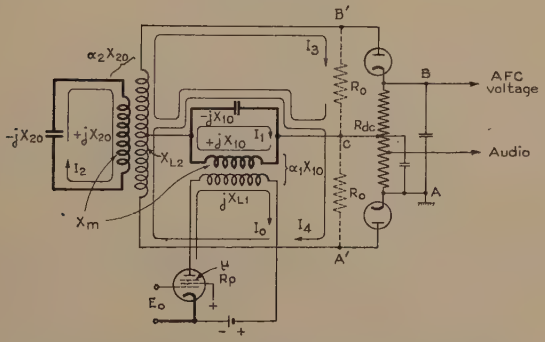


Fig. 5—Complete automatic-frequency-control discriminator circuit.

We find from this set of equations

$$I_2 = \frac{-\mu E_0}{D}\,\frac{\alpha_1 X_m}{X_{20}} = \frac{E_2}{-jX_{20}} \tag{11}$$

$$I_3 = -\frac{\mu E_0}{D}\,\frac{\alpha_1 X_{10}}{R_0}\left[p_2 + j2\delta + j\tfrac{1}{2}\alpha_2 k\sqrt{\frac{X_{20}}{X_{10}}}\right] \tag{12}$$

$$I_4 = +\frac{\mu E_0}{D}\,\frac{\alpha_1 X_{10}}{R_0}\left[p_2 + j2\delta - j\tfrac{1}{2}\alpha_2 k\sqrt{\frac{X_{20}}{X_{10}}}\right] \tag{13}$$

$$D = R_p[(p_1 + j2\delta)(p_2 + j2\delta) + k^2]\,. \tag{14}$$

p_1 and p_2 are the power factors of the primary and secondary circuit respectively. From the foregoing calculation, they become

$$p_1 = \frac{1}{Q_1} + \alpha_1^2 \frac{X_{10}}{R_p} + \frac{X_{10}}{\frac{1}{2}R_0} \tag{15}$$

$$p_2 = \frac{1}{Q_2} + \alpha_2^2 \frac{X_{20}}{2R_0} \tag{16}$$

$$k^2 = \frac{X_m^2}{X_{10}X_{20}} = \text{coefficient of coupling between primary and secondary circuits} \tag{17}$$

$$A = \tfrac{1}{2}\alpha_2 k \sqrt{\frac{X_{20}}{X_{10}}}. \tag{18}$$

Assuming, for the moment, the transformer ratios α_1 and α_2 equal to unity, we note a complete identity between (11) and (3). Comparing now the power factors (equations (15) and (16)) with those for the standard intermediate-frequency transformer (equations (4) and (5)), we note that a term $X_{10}/\frac{1}{2}R_0$ has been added in the expression for p_1, while p_2 now has a loading term $X_{20}/2R_0$. This means that the primary circuit is loaded by a parallel load of $\frac{1}{2}R_0 = \frac{1}{8}R_{dc}$ (both R_0's in parallel) and the secondary by a load $2R_0 = \frac{1}{2}R_{dc}$ (both R_0's in series). This relatively heavy load on both the primary and secondary circuit is responsible for the poor selectivity of the automatic-frequency-control transformer.

The similarity of (11), (14), (15), and (16) to (3), (4), and (5) confirms the statement made above regarding the close relationship between the automatic-frequency-control and the standard intermediate-frequency transformer.

III. Automatic-Frequency-Control Discriminator Performance

The radio-frequency voltages, which appear on the diodes, shown in Figs. 3 and 5, are I_3R_0 and I_4R_0. These voltages vary with frequency. We shall now study their behavior in more detail.

A. Vector Diagram for the Diode Voltages

We have, from (12) and (13),

$$\frac{I_3R_0}{E_0} = \frac{E_3}{E_0} = -\frac{\mu}{R_p}\alpha_1 X_{10} \frac{p_2 + j(2\delta + A)}{(p_1 + j2\delta)(p_2 + \delta 2\delta) + k^2} \tag{19}$$

$$\frac{I_4R_0}{E_0} = \frac{E_4}{E_0} = +\frac{\mu}{R_p}\alpha_1 X_{10} \frac{p_2 + j(2\delta - A)}{(p_1 + j2\delta)(p_2 + j2\delta) + k^2}. \tag{20}$$

If all of the magnitudes on the right-hand side are known, E_3/E_0 can be computed versus 2δ. More convenient for this computation, however, is a vector diagram, which we shall now derive. We write

$$\frac{E_3}{E_0} = -j\frac{\mu}{R_p}\,\frac{1}{2}\alpha_1\alpha_2\,\frac{\sqrt{X_{10}X_{20}}}{k}\,\frac{N_3}{M} \tag{21}$$

where,

$$N_3 = -j\frac{p_2}{A} + \frac{1}{1+jy} = |N_3|\,\epsilon^{j\phi} \tag{22}$$

$$M = \frac{p_2^2}{k^2}\left(\frac{p_1}{p_2} + jy\right) + \frac{1}{1+jy} = |M|\,\epsilon^{j\psi} \tag{23}$$

$$y = \frac{2\delta}{p_2} = \frac{2}{p_2}\,\frac{\omega-\omega_0}{\omega_0}. \tag{24}$$

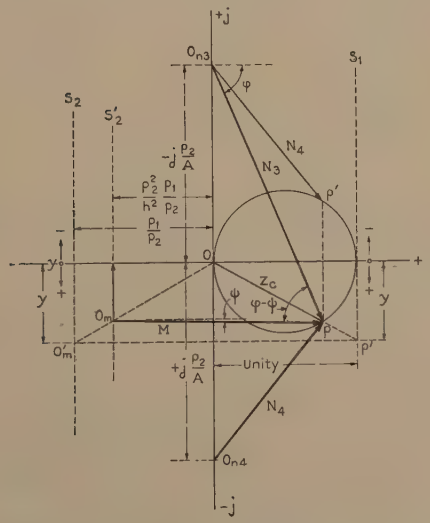


Fig. 6—Vector diagram for determining E_3 and E_4.

In these equations we have introduced a new variable y which is proportional to 2δ.

The complex locus curve representing the term

$$Z_c = \frac{1}{1+jy}$$

becomes a circle for variable y. The diameter of this circle is unity. The vectors N_3 and M can be represented in a single vector diagram.

We see from Fig. 6 that

$$N_3 = \overline{O_{n3}P} = -j\frac{p_2}{A} + Z_c$$

$$M = \overline{O_mP} = +j\frac{p_2^2}{k^2}y + \frac{p_2^2}{k^2}\frac{p_1}{p_2} + Z_c.$$

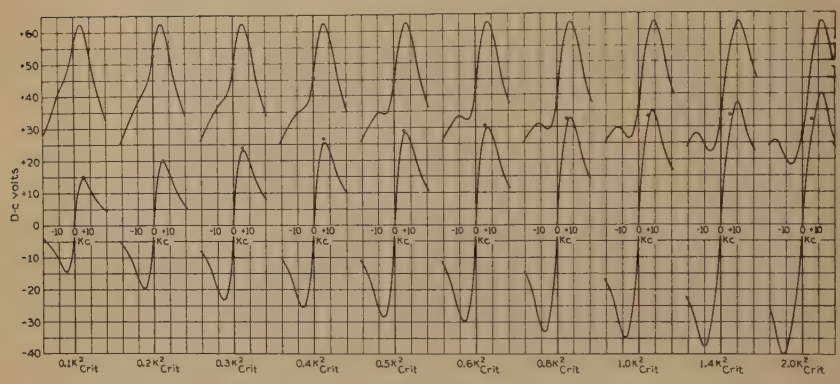


Fig. 7—Upper row: E_1/E_0 versus frequency.
Lower row: $\dfrac{|E_4| - |E_3|}{E_0}$ versus frequency.

All curves are for various values of coupling between primary and secondary. Small circles near maxima of lower row indicate values resulting from the approximate relation (30). The data for this set of curves are

$p_1 = 50.10^{-3}$
$p_2 = 25.10^{-3}$
$\alpha_1 = 0.415$
$\alpha_2 = 1.0$

$X_{10} = 5840$ ohms
$X_{20} = 7600$ ohms
$p_1/p_2 = 2.0$

For E_4/E_0 we simply have to write $-A$ instead of $+A$; hence,

$$\frac{E_4}{E_0} = -j\frac{\mu}{R_p}\frac{1}{2}\alpha_1\alpha_2\frac{\sqrt{X_{10}X_{20}}}{k}\frac{N_4}{M} \tag{25}$$

where,

$$N_4 = \overline{O_{n4}P} = +\frac{p_2}{A} + Z_c. \tag{26}$$

Thus, the vectors N_3, N_4, and M can be determined in both amplitude and phase from the vector diagram of Fig. 6. As seen from (21), the phase angle between E_3 and E_0 equals

$$90° + \phi - \psi.$$

The angle $\phi - \psi$ is equal to the angle $O_{n3} - P - O_m$; hence this angle may be read directly from the diagram without additional calculation.

For instance, to find E_3/E_0 from the vector diagram, first compute p_2/A, p_1/p_2, $p_1p_2^2/p_2k^2$. Draw scales S_1 and S_2 parallel to the j axis, with distances from the j axis equal to 1 and p_1/p_2, respectively. Calibrate scales S_1 and S_2 in terms of y (or frequency) Draw a circle having its diameter equal to 1. Draw a line OO_m' and a line OP'. Make $OO_{n3}=p_2/A$. Draw S_2' parallel to the j axis for a distance $p_1p_2^2/p_2k^2$. Measure N_3 and M, and $\phi-\psi$. Compute E_3/E_4 from (21).

A series of E_3/E_0 curves for various values of coupling is shown in Fig. 7. The operating points are marked. These curves have been computed by means of the above vector diagram.

B. Sensitivity Factor

Referring to the vector diagram, Fig. 4, it is seen that the automatic-frequency-control direct voltage, between $A-B$ in Fig. 9, is proportional to $|E_4|-|E_3|$, if the rectifier is linear. From (19) and (20)

$$\frac{|E_4|-|E_3|}{E_0}=\frac{\mu}{D}\alpha_1X_{10}\left[\sqrt{p_2^2+(2\delta-A)^2}-\sqrt{p_2^2+(2\delta+A)^2}\right]. \quad (27)$$

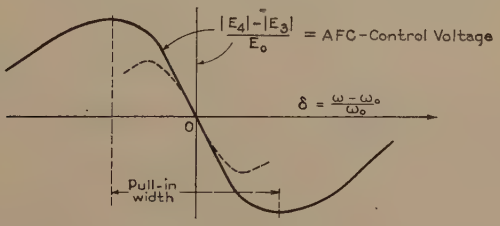


Fig. 8

The voltage difference $|E_4|-|E_3|$ will vary with frequency as shown in Fig. 8. The variation is fastest at $\delta=0$. The equation of the tangent of the curve in that point is

$$\frac{|E_4|-|E_3|}{E_0}=\left[\frac{\mu}{R_p}\,\frac{\alpha_1\alpha_2k\sqrt{X_{10}X_{20}}}{p_1p_2+k^2}\,\frac{1}{\sqrt{p_2^2+A^2}}\right]2\delta. \quad (28)$$

The bracketed term in (28) is the automatic-frequency-control sensitivity factor, which represents the slope of the curve, Fig. 8, at $\delta=0$.

The automatic-frequency-control direct voltage is obtained by multiplying with $\sqrt{2}$ (assuming 100 per cent rectification efficiency) on the right-hand side; thus for an intermediate frequency of 465 kilocycles

$$\frac{\text{automatic-frequency-control direct voltage}}{E_0 \text{ (root-mean-square volts)}} \text{ per 1-kilocycle deviation}$$

$$= \frac{\mu}{R_p} \frac{\alpha_1 \alpha_2 k \sqrt{X_{10} X_{20}}}{p_1 p_2 + k^2} \frac{2\sqrt{2}}{\sqrt{p_2^2 + A^2}} \frac{1}{465}. \tag{29}$$

Curves of automatic-frequency-control voltage versus frequency are shown in Fig. 7.

C. Peak Separation

Referring to Fig. 8, the distance between the two maxima may be called "peak separation." This width may be measured in per cent of intermediate frequency or in cycles. Over this range the automatic-frequency-control discriminator will provide a direct voltage whose amplitude and sign correspond to the frequency deviation from the nominal intermediate frequency.

It would be of interest to the designer to know the exact location of the maxima and minima of the function $(|E_3| - |E_4|)/E_0$ shown in (27). Unfortunately, this function is too complicated to permit the computation of the maxima in the usual manner.

Because amplitude and location of the maxima cannot be computed directly, the only remaining alternative is to determine the function E_3/E_0 from the vector diagram, Fig. 6, from which the maxima of (19), (20), and (27) can be found.

If the coupling between primary and secondary is small, an approximate formula may be used. In this case p_2/A and p_2^2/k^2 are much larger than 1 and the vectors $\overline{O_{n3}O}$ and $\overline{O_mO}$ become much greater than 1. Consequently, $|N_3| - |N_4'|$ (which is equal to $|N_3| - |N_4|$) becomes nearly equal to the secant PP'. But PP' assumes a maximum if $y = 1$. Hence the maximum of (27) occurs if $y = 1$ or

$$y = \frac{2\delta}{p_2} = \frac{2}{p_2} \frac{\omega - \omega_0}{\omega_0} = 1.$$

This means

$$2\delta = p_2 = \text{"peak separation" for small } k. \tag{30}$$

In the lower row of figures of Fig. 7, one will find small circles near the maximum value of the $(|E_3| - |E_4|)/E_0$ curves. These circles represent the location of the maximum according to the approximate formula (30). It is seen that for $k < 0.75\, k_{crit}$ the approximative results are quite satisfactory. For greater k, the actual peak separation is greater than the value computed from the approximate formula.

The curves which are shown in Fig. 7 are all computed for a constant grid-input voltage, E_0. In a radio set, however, due to the selectivity of the preceding stages, E_0 varies with frequency. This causes a change in the automatic-frequency-control voltage characteristic as shown dotted in Fig. 8. In other words, the "over-all" peak separation is smaller than that of the discriminator alone. (The slope of the automatic-frequency-control voltage characteristic at $\delta = 0$, however (equation (28)) is not affected by this fact.) On account of this item of uncertainty, the above approximate formula is satisfactory for estimate purposes.

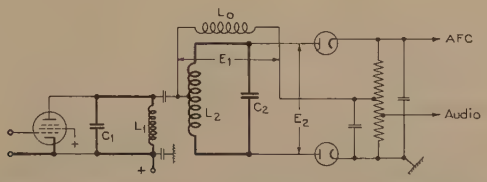


Fig. 9—Original automatic-frequency-control discriminator circuit.

The width for peak separation is usually not the same as the "pull-in width." The actual "pull-in width" is a function of control-tube characteristics as well as of the discriminator circuit and the selectivity of the preceding stages.

IV. Modifications of the Fundamental Automatic-Frequency-Control Circuit

The next step will be to show that the formulas derived above apply to the circuit which is now in practical use. This circuit is shown in Fig. 9. E_1 and E_2 are connected in series by using a choke L_0 and a coupling capacitor C_0.

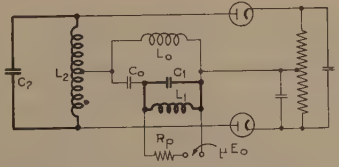


Fig. 10—Fig. 9, redrawn.

A. Choke of Small Capacitance

First consider a choke (L_0) whose distributed capacitance is negligibly small. The circuit of Fig. 9 can be redrawn into that of Fig. 10. The primary circuit alone appears in Fig. 11. The voltage E_1 lies across L_0; consequently, X_{L0} must be used in place of X_{10} in the previously

derived formulas. Since the complete circuit must resonate at the intermediate frequency, it follows that the circuit L_1C_1 must be capacitively reactive. The resulting impedance of the L_1C_1 circuit which appears between the terminals 1 and 2 is

$$R_1\left(\frac{X_{c1}}{X_{L1}-X_{c1}}\right)^2 - j\frac{X_{L1}X_{c1}}{X_{L1}-X_{c1}} = R'_{c1} - jX'_{c1}. \tag{31}$$

(This formula is invalid if $X_{L1}=X_{c1}$.)

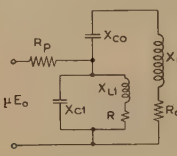


Fig. 11

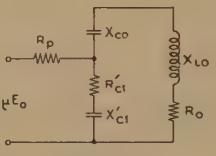


Fig. 12

We can now draw Fig. 12 as the equivalent circuit of Fig. 11. The resonant condition is $X_{L0}=X_{c0}+X'_{c1}$, hence

$$\frac{X_{L0}-X_{c0}}{X_{L1}} = \frac{X_{c1}}{X_{L1}-X_{c1}}.$$

The capacitor combination C_0C_1 acts like an autotransformer; the step-down ratio is

$$\alpha_1 = \frac{X'_{c1}}{X_{c0}+X_{c1}} = \frac{X_{L0}-X_{c0}}{X_{L0}} = 1 - \frac{X_{c0}}{X_{L0}}.$$

Next, we have to find the resulting p_1. We have from Fig. 12

$$\frac{1}{Q_{\text{res}}} = \frac{1}{X_{L0}}(R_0+R'_{c1}) = \frac{1}{X_{L0}}\left[R_0 + R_1\left(\frac{X_{c1}}{X_{L1}-X_{c1}}\right)^2\right]$$

$$= \frac{1}{X_{L0}}\left[R_0 + R_1\left(\frac{X_{L0}-X_{c0}}{X_{L1}}\right)^2\right] = \frac{1}{Q_0} + \alpha_1^2\frac{X_{L0}}{X_{L1}}\frac{1}{Q_1},$$

where Q_0 and Q_1 are the quality factors of the choke L_0 and the primary inductance L_1, respectively. Now p_1 becomes

$$p_1 = \frac{1}{Q_0} + \alpha_1^2\frac{X_{L0}}{X_{L1}}\frac{1}{Q_1} + \alpha_1^2\frac{X_{L0}}{R_p} + \frac{X_{L0}}{\frac{1}{2}R_0}.$$

For the secondary circuit we have

$$\alpha_2 = 1$$

$$p_2 = \frac{1}{Q_2} + \frac{X_{20}}{2R_0}.$$

B. Choke of Medium Capacitance

This refers to a choke whose resonant frequency is considerably higher than the intermediate frequency, but whose distributed capacitance is of fairly large size and cannot be neglected. In Fig. 13 the distributed capacitance is shown as a lumped capacitance having the reactance $-jX_{cp}$.

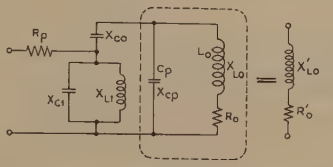


Fig. 13

This case can be readily converted into the one just considered if we apply (31) to the network X_{L0}, R_0, X_{cp}. This network is equivalent to a series impedance $R'_0 + j\,X_{L0}'$ which is found to be

$$R'_0 + j\,X_{L0}' = R_0\left(\frac{X_{cp}}{X_{L0} - X_{cp}}\right)^2 - j\,\frac{X_{L0}\,X_{cp}}{X_{L0} - X_{cp}}\,.$$

If we denote $\omega_r = 1/\sqrt{L_0 C_p}$ = the resonant frequency of the choke, we obtain from the above (with ω_0 = intermediate frequency)

$$X_{L0}' = X_{L0}\,\frac{\omega_r^2}{\omega_r^2 - \omega_0^2} \tag{32}$$

$$Q_0' = \frac{X_{L0}'}{R_0'} = Q_0\left(1 - \frac{\omega_0}{\omega_r}\right)^2. \tag{33}$$

By means of (32) and (33), this case can be converted into the one just considered previously.

C. Choke Resonating at the Intermediate Frequency

If the choke happens to resonate close to or at the intermediate frequency, the treatment used in the two foregoing sections cannot be used, because (31) does not hold if either L_1C_1 or L_0C_p resonates at the operating frequency. If L_0C_p resonates at the intermediate frequency, then L_1C_1 also must resonate at that frequency. The resonant impedance of L_1C_1 is very high. X_{c0} becomes small in comparison with this impedance. Hence α_1 becomes equal to 1. The resulting X_{10} becomes

$$X_{10} = \frac{X_{L0}X_{L1}}{X_{L0} + X_{L1}}$$

and for the resulting Q_{res} we find

$$Q_{res} = \frac{(X_{L0} + X_{0L})Q_1Q_0}{Q_1X_{L1} + Q_0X_{L0}}.$$

Measurements were made on an automatic-frequency-control transformer having the following data:

$L_1 = 0$ 90 millihenrys	$X_{L1} = 2630$ ohms	$Q_1 = 146$
$L_2 = 2.47$ millihenrys	$X_{20} = 7200$ ohms	$Q_2 = 106$
$C_0 = 85$ micromicrofarads	$X_{c0} = 4000$ ohms	$Q_0 = 40$
$L_0 = 15$ millihenrys	$X_{L0} = 43800$ ohms	

The resonant frequency of the choke was 465 kilocycles. The following results were obtained:

TABLE I

Coupling	Direct Voltage		
	Gain	AFC at 1 kc off	
$\sqrt{0.6}\ k_{crit}$	106	19.4	calculated
$\sqrt{0.8}\ k_{crit}$	100	18.5	calculated
$1.0\ k_{crit}$	95	17.6	calculated
about k_{crit}	90	19	measured

D. Double-Ended Primary—Single-Ended Secondary

The circuit shown in Fig. 14 is particularly suitable if the automatic frequency control and the signal channel are to be separated. It is not difficult to compute the currents in this type of network and correlate them with those previously derived. The separate signal channel, shown dotted in Fig. 14, will be omitted for the time being.

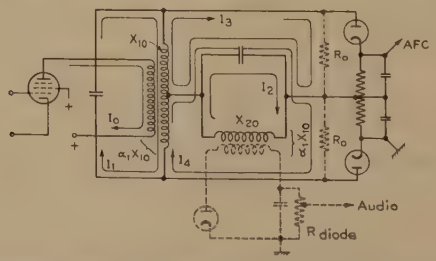


Fig. 14—Modified automatic-frequency-control discriminator circuit.

The nomenclature is illustrated in Fig. 14. If the same procedure is followed as in Section II above, the results are

$$I_2 = -\frac{\mu E_0}{D}\,\frac{\alpha_1 X_m}{X_{20}} = \frac{E_2}{-jX_{20}} \qquad (11a)$$

$$I_3 = + \frac{\mu E_0}{D} \frac{\frac{1}{2}\alpha_1 X_{10}}{R_0} (p_2 + j(2\delta + A)) \tag{12a}$$

$$I_4 = + \frac{\mu E_0}{D} \frac{\frac{1}{2}\alpha_1 X_{10}}{R_0} (p_2 + j(2\delta - A)) \tag{13a}$$

$$A = 2k \sqrt{\frac{X_{20}}{X_{10}}} \tag{18a}$$

$$D = R_p[(p_1 + j2\delta)(p_2 + j2\delta) + k^2] \tag{14a}$$

$$p_1 = \frac{1}{Q_1} + \frac{\alpha_1^2 X_{10}}{R_p} + \frac{X_{10}}{2R_0} \tag{15a}$$

$$p_2 = \frac{1}{Q_2} + \frac{X_{20}}{\frac{1}{2}R_0}. \tag{16a}$$

For the construction of the vector diagram

$$\frac{E_3}{E_0} = + j \frac{\mu}{R_p} \alpha_1 \frac{\sqrt{X_{10}X_{20}}}{k} \frac{N_3}{M}. \tag{21a}$$

$$\frac{E_4}{E_0} = - j \frac{\mu}{R_p} \alpha_1 \frac{\sqrt{X_{10}X_{20}}}{k} \frac{N_4}{M}. \tag{25a}$$

For the automatic-frequency-control sensitivity factor we have

$$\frac{|E_4| - |E_3|}{E_0} = \left[\frac{\mu}{R_p} \frac{2\alpha_1 k \sqrt{X_{10}X_{20}}}{p_1 p_2 + k^2} \frac{1}{\sqrt{p_2^2 + A^2}} \right] 2\delta .$$

An automatic-frequency-control transformer of this type was tested, which had data as follows:

$L_1 = 2.31$ millihenrys $\quad X_{10} = 6750$ ohms $\quad Q_1 = 110$
$L_2 = 0.914$ millihenry $\quad X_{20} = 2670$ ohms $\quad Q_2 = 140$
$L_{1m} = 1.16$ millihenrys $\quad \alpha_1 = 0.50$
$R_0 = \frac{1}{2} \cdot 1{,}000{,}000 = 500{,}000$ ohms

A separate signal diode was used, coupled to the secondary circuit. For this channel

$L_{2m} = 0.459$ millihenrys $\quad \alpha_2' = 0.50$
R_0 diode $= \frac{1}{2} \cdot 300{,}000 = 150{,}000$ ohms $\quad R_p = 800{,}000$ ohms.
$\mu = 1000$;

The following results were obtained:

TABLE II

Coupling	Direct Voltage			
	ACA	Signal Gain	AFC at 1 kc off	
$\sqrt{0.8}\ k_{crit}$	2.75	47.5	25.3	calculated
$\sqrt{1.0}\ k_{crit}$	2.40	46.9	23.2	calculated
$\sqrt{1.2}\ k_{crit}$	2.13	46.5	22.1	calculated
about k_{crit}	2.25	50	23.5	measured

V. Performance of the Automatic-Frequency-Control Discriminator Circuit as a Signal-Diode Transformer

The audio-frequency signal can be taken off from the automatic-frequency-control circuit in two different ways: either between points A and C in Fig. 5, or from a separate signal diode as shown dotted in Fig. 14.

A. Common Signal and Automatic-Frequency-Control Channels

The voltage E_4 is assymetrical both in amplitude and in phase with reference to the intermediate-frequency carrier. (See Fig. 7.) In the circuit of Fig. 5, the points A and B are connected by a large capacitor. We must, therefore, investigate not only the envelope of E_4 but also that of E_3.

We consider an amplitude-modulated signal and compute amplitude and phase of carrier and side bands by means of the vector diagram Fig. 6. We find for the transformer (data given above under "choke resonating at intermediate frequency"):

TABLE III

		$f=1000$		$f=2500$		$f=5000$		$f=10{,}000$	
		E_3/E_0	$\phi-\psi$	E_3/E_0	$\phi-\psi$	E_3/E_0	$\phi-\psi$	E_3/E_0	$\phi-\psi$
E_3	Upper side band	106	−139.5	120.5	−147.5	141.4	−166	131	−208
	Carrier	97.3	−135.5	97.3	−135.5	97.3	−135.5	97.3	−135.5
	Lower side band	88	−132	78.5	−129	66.8	−126.5	57	−120.5
		E_4/E_0	$\phi-\psi$	E_4/E_0	$\phi-\psi$	E_4/E_0	$\phi-\psi$	E_4/E_0	$\phi-\psi$
E_4	Upper side band	88	−48	78.5	−51	66.8	−53.5	57	−59.5
	Carrier	97.3	−44.5	97.3	−44.5	97.3	−44.5	97.3	−44.5
	Lower side band	106	−41.5	120.5	−32.5	141.4	−14.0	131	+28

Applying known methods of radio-frequency-envelope analysis,[9] one obtains the two ellipses shown in Fig. 15 as locus curves of the modulation vector. Plotting the radio-frequency envelope from these ellipses, we get the curves shown in Fig. 16.

[9] H. Roder, "Some notes on demodulation," Proc. I.R.E., vol. 20, pp. 1946–1961; December, (1932).

We note from Fig. 15 that the instantaneous values of the envelopes of E_3 and E_4 are exactly equal at every instant during the audio-frequency cycle. Consequently, the rectified voltages, CA and CB (Fig. 5), are equal at every instant during the audio-frequency cycle. Points A and B are always at an equal potential. This statement was

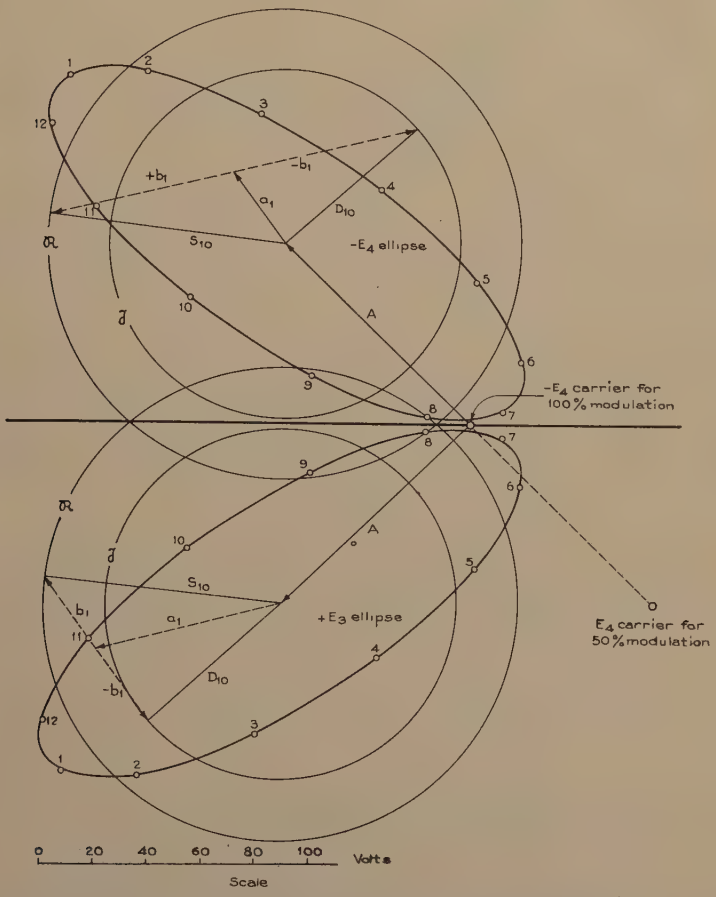


Fig. 15—Modulation vector diagram for the modulated signal. 5000-cycle modulation frequency, data in Table III.

checked experimentally and found to be correct. Both rectifiers work fully independently of each other. The envelope, i.e., the wave shape of the audio-frequency voltage, is given by the ellipse for E_3 in Fig. 15.

Fig. 16 shows the resulting E_3 envelope plotted versus time. For 5000 cycles and 100 per cent modulation the audio-frequency distortion is quite appreciable, but at 50 per cent modulation it appears to

be very small. The reason for distortion is not so much that the upper and lower side bands are of different amplitude, but rests in the fact that the relative phase displacements of the side bands from the carrier are unequal. This can be noted from the above tabulation. At 10,000

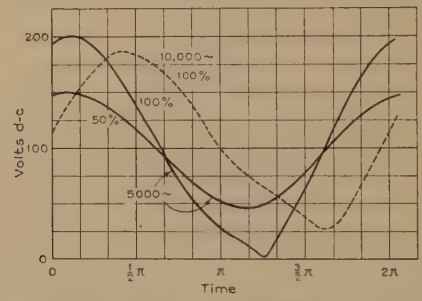


Fig. 16—Radio-frequency envelopes for 5000- and 10,000-cycle modulation frequency, for data in Table III.

cycles and 100 per cent modulation, the envelope distortion increases rapidly However, at frequencies from about 3000 cycles up, the modulation depth is usually rather small. Thus, in practice the above-described type of envelope distortion is generally not a matter of much consequence, except in high-fidelity receivers.

B. Separate Automatic-Frequency-Control and Signal Channels

The automatic-frequency-control and the signal channel can be separated in case an extra diode is available. This arrangement has a number of advantages, such as more adjacent-channel attenuation,

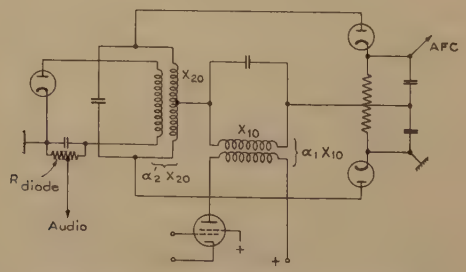


Fig. 17—Discriminator circuit with separate signal diode.

more automatic-frequency-control sensitivity, and less envelope distortion. The signal gain, however, is somewhat less.

Two possible arrangements are shown in Fig. 14 (dotted) and in Fig. 17. In both cases the diode is inductively coupled to the tuned circuit.

We have for Fig. 14

$$\alpha_1 = 1,$$

$$p_1 = \frac{1}{Q_1} + \frac{X_{10}}{2R_0} + \alpha_1^2 \frac{X_{10}}{R_p},$$

$$p_2 = \frac{1}{Q_2} + \frac{X_{20}}{\frac{1}{2}R_0} + \alpha_2'^2 \frac{X_{20}}{\frac{1}{2}R_{\text{diode}}}.$$

The audio-frequency-channel gain, computed from (3), is

$$\text{gain} = \frac{\mu}{R_p} \frac{k\sqrt{X_{10}X_{20}}}{p_1p_2 + k^2} \alpha_1\alpha_2'.$$

Adjacent-channel attenuation and band width can be computed from (6). The automatic-frequency-control sensitivity follows from (28a).

For Fig. 17 we have

$$\alpha_2 = 1$$

$$p_1 = \frac{1}{Q_1} + \frac{X_{10}}{\frac{1}{2}R_0} + \alpha_1^2 \frac{X_{10}}{R_p}$$

$$p_2 = \frac{1}{Q_2} + \frac{X_{20}}{2R_0} + \alpha_2'^2 \frac{X_{20}}{\frac{1}{2}R_{\text{diode}}}.$$

The audio-frequency-channel gain is (equation (3))

$$\text{gain} = \frac{\mu}{R_p} \frac{k\sqrt{X_{10}X_{20}}}{p_1p_2 + k^2} \alpha_1\alpha_2'.$$

The automatic-frequency-control sensitivity is given by (29).

The circuits shown in Figs. 14 and 17 have a number of advantages.

1. The automatic-frequency-control action does not interfere with the signal channel. There is no nonlinear envelope distortion for the signal envelope.
2. The automatic-frequency-control-diode load resistors can be appreciably increased in size. This results in reduced load on the tuned circuits and compensates for the added load due to the signal diode.
3. For the signal, the circuit behaves like an ordinary diode-feeding transformer. The obtainable adjacent-channel attenuation is from 2 to 3. With common automatic-frequency-control and signal channels, the adjacent-channel attenuation is in the order of from 1 to 1.5.

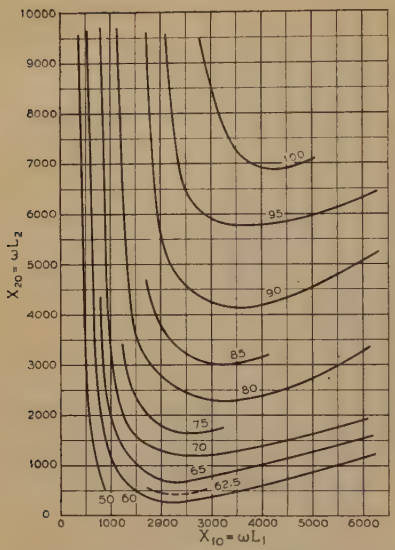


Fig. 18—Curves of constant gain. "Gain" expressed in direct-voltage output for a constant 1-volt root-mean-square input.

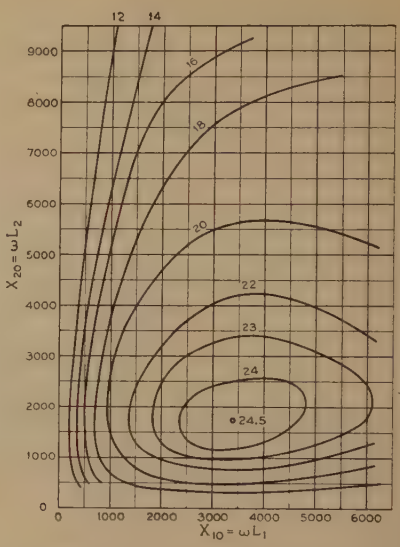


Fig. 19—Curves of constant automatic-frequency-control voltage. Direct-voltage output for a 1-volt root-mean-square input and for 1-kilocycle detuning.

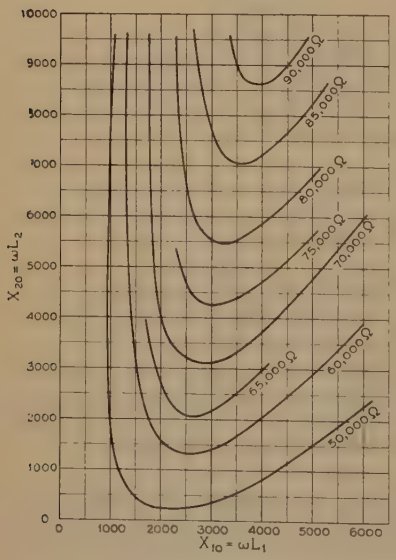


Fig. 20—Curves of constant input impedance.

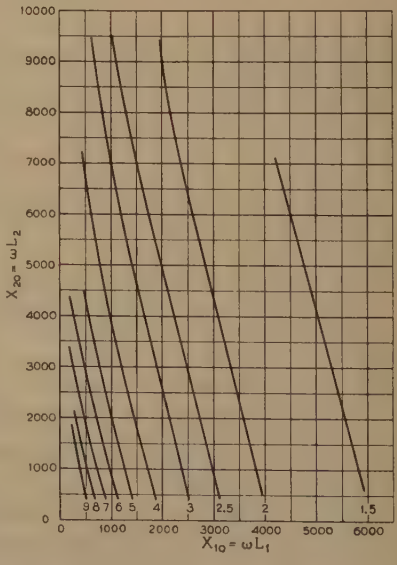


Fig. 21—Curves of constant adjacent-channel attenuation for separate automatic-frequency-control and signal channels.

The only disadvantage is that the audio-frequency-channel gain is less than it is in case of common automatic-frequency-control and signal channels.

VI. Design Curves

For the circuit shown in Fig. 17, but without the extra signal diode, a number of "design curves" have been worked out. Figs. 18 to 21 show $\omega_0 L_2$ plotted versus $\omega_0 L_1$ with signal gain, automatic-frequency-control voltage per kilocycle, input impedance (as seen from the driver tube), and adjacent-channel attenuation, as parameters. The following circuit data have been assumed:

$Q_1 = 125$ $\quad R_p = 750{,}000$ ohms

$Q_2 = 111$ $\quad \mu = 1000$

frequency = 465 kilocycles

direct-current load resistor per diode = 500,000 ohms

The coupling was in all cases assumed to be undercritical; namely, $k^2 = 0.7\ k^2_{crit} = 0.35\ (p_1^2 + p_2^2)$. The transformer ratios α_1 and α_2 have been put equal to 1. For computing gain and automatic-frequency-control voltages a rectifier efficiency of 85 per cent was assumed.

The data presented in Figs. 18 to 21 apply to the case in which the signal- and the automatic-frequency-control channel are common. The adjacent-channel attenuation cannot be computed in this case with any reasonably degree of accuracy. The adjacent-channel-attenuation data given in Fig. 21 would result if a separate signal diode were coupled to the secondary circuit (provided this diode would not represent an additional load). The actual adjacent-channel attenuation will be less than the values given, probably about 50 per cent. The adjacent-channel-attenuation chart can therefore merely serve as an indication as to the direction in which the adjacent-channel attenuation will change if X_{10} or X_{20} is varied.

Acknowledgment

Mr. D. F. Vrooman of the Radio Receiver Engineering Section, Bridgeport Works, General Electric Company, has kindly supplied all the experimental data to which reference is made in this paper.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

GRID-CURRENT FLOW AS A FACTOR IN THE DESIGN OF VACUUM-TUBE POWER AMPLIFIERS*

By

W. L. Everitt and Karl Spangenberg
(Ohio State University, Columbus, Ohio)

Summary—*In the design of class C amplifiers analytical methods have been proposed which give good results from the standpoint of plate-circuit operating condition. Similar methods have not been available for the determination of grid-circuit conditions. It is the purpose of this paper to describe a method by which this result can be obtained.*

An analytical determination of grid-circuit operating conditions requires a knowledge of grid current as a function of the grid and plate voltages. Since good approximations for the space current in a triode are available, an expression for grid current can be had if the ratio of plate-to-grid current as a function of plate and grid voltage can be determined. Theoretical considerations show that the ratio of plate current to grid current should be a function of the ratio of plate voltage to grid voltage. The nature of this functional relation is determined experimentally and a comparatively simple and accurate expression for grid current in cases in which there is only a small amount of secondary emission is thus obtained.

The design procedure for class C amplifiers is based upon the allowable grid and plate dissipation. The method of determining optimum operating conditions on the basis of plate-circuit analysis alone is modified to include the analytical determination of the corresponding grid-driving power and grid loss. This procedure results in the determination of the operating conditions which will give the highest output which can be obtained without exceeding the allowable grid or plate dissipation. The design procedure is facilitated by the use of universally applicable functions presented in graphical form. Experimental results are in good agreement with the theory developed.

Introduction

IN PREVIOUS publications[1,2,3] a method making use of reasonable approximations has been developed for the computation of the operating conditions of class C and class B radio-frequency amplifiers which will produce predetermined conditions of loss, output, and efficiency in the plate circuit. In the criteria which were there set up no attention was paid to the possible losses which might exist in the grid circuit. In the present paper it will be shown how the character-

* Decimal classification: R355.7. Original manuscript received by the Institute, August 30, 1937. Presented before Pacific Coast Meeting, Spokane, Washington, September 1, 1937.

[1] W. L. Everitt, "Optimum operating conditions for class C amplifiers," Proc. I.R.E., vol. 22, pp. 152–176; February, (1934).

[2] W. L. Everitt, "Optimum operating conditions for class B radio-frequency amplifiers," Proc. I.R.E., vol. 24, pp. 305–315; February, (1936).

[3] W. L. Everitt, "Communication Engineering," Second Edition, chapter XVII. McGraw-Hill Book Company, (1937).

istics of the grid circuit may also be computed and a modification of the "optimum condition" is proposed. It should be recognized that an "optimum condition" must depend on a number of factors and so must be defined for any type of operation. The design formulas here offered as well as those in the first paper[1] can also be used under other conditions of operation.

The losses in the grid circuit of a tube are divided into two parts: (1) those in the grid-bias device, and (2) those due to the electronic bombardment of the grid which result in the heating of the grid. These total losses must be supplied by the driving source. This gives rise to the equation

$$W_d = E_g I_{g1}/2 = E_c I_{g0} + W_g \tag{1}$$

where,

W_d is the total power supplied to the grid circuit
E_g is the maximum value of the sinusoidal voltage driving the grid
I_{g1} is the maximum value of the fundamental component of the grid current
E_c is the absolute value of the negative bias
I_{g0} is the average or direct-current component of the grid current
W_g is the power dissipated on the grid.

Spitzer[4] obtained an experimental relation for W_g. This relation is for cases in which the grid-driving voltage alone is varied,

$$W_g = A I_{g0} 1.34 \tag{2}$$

where A is a constant which must be determined for each tube and each operating condition. This expression is not convenient for the predetermination of grid losses as no method has been given for the computation of I_{g0} or A. They are essentially dependent variables whose values are functions of the tube used, the circuit constants, and the applied voltages.

Terman[5] and Thomas[6] have given a simple approximate relation for the determination of the grid-driving power in the equation

$$W_d = E_g I_{g0}. \tag{3}$$

While the values of (3) are readily measured by meters, this equation has not been used in the predetermination of grid-driving power because of the difficulty in the computation of I_{g0}.

[4] E. E. Spitzer, "Grid losses in power amplifiers," Proc. I.R.E., vol. 17, pp. 985–1005; June, (1929).
[5] F. E. Terman, "Radio Engineering," First Edition, p. 234. McGraw-Hill Book Company, (1932).
[6] H. P. Thomas, "Determination of grid driving power in radio-frequency amplifiers," Proc. I.R.E., vol. 21, pp. 1134–1141; August, (1933).

The point-by-point analysis given by Prince[7] and developed by Mouromtseff and Kozanowski[8] is applicable to the analysis of the grid circuit on the same basis as the plate circuit. This method is lengthy and does not lend itself to a quick determination of optimum conditions as was discussed in the original paper.[1]

The Grid-Current Equation

One of the difficulties in the way of the accurate computation of grid losses has been the presence of secondary emission from the grid. When the grid voltage is positive, but less than the plate voltage, secondary electrons may be emitted from the grid and attracted to the plate thus reducing the net grid current. This results in a reduction of the losses in the grid circuit, which by itself is desirable, but at the same time it frequently produces an erratic behavior which is most undesirable. There seems to be a tendency on the part of tube manufacturers to reduce the amount of this secondary emission in tubes in order to improve the stability of amplifiers. With such tubes reasonable accuracy may be expected in the predetermination of grid losses and driving power. For that reason this paper will be confined to an analysis of the action of tubes in which secondary emission plays a minor factor. Experimental tests were made on tubes with tantalum electrodes where the omission of this factor is warranted.

Grid current is of interest when the grid voltage is positive. This occurs during only a fraction of the cycle in the operation of a class B or class C amplifier. The path of an electron in a tube is determined by the geometrical configuration of the lines of force within the tube. These lines of force are determined by the geometry of the tube, the grid and plate potentials, and the distribution of the space charge. For given values of these parameters a certain proportion of the space current (total current leaving the filament) will be diverted to the grid and the remainder will go to the plate of a triode. If both the plate and grid voltages are changed by a given ratio, the number of lines of force will change, but their configuration will not. Hence the general paths of the electrons will remain the same and *the proportion of the total current passing to each electrode will not change.* This shows that, if secondary emission may be neglected, the following equation can be written

$$i_p/i_g = f(e_p/e_g) \tag{4}$$

where i_p and i_g are respectively the instantaneous values of the plate

[7] D. C. Prince, "Vacuum tubes as power amplifiers," Proc. I.R.E., vol. 11, pp. 275–332; June; pp. 405–458; August; and pp. 527–550; October, (1923).

[8] Mouromtseff and Kozanowski, "Analysis of the operation of vacuum tubes as class C amplifiers," Proc. I.R.E., vol. 23, pp. 752–778; July, (1935).

and grid current and e_p and e_g the instantaneous values of plate and grid voltage. In order to check the form of the function of (4) the

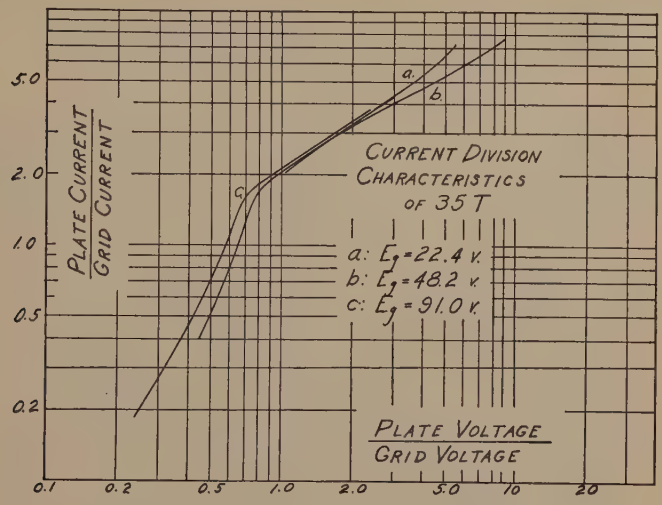


Fig. 1—i_p/i_g as a function of e_p/e_g.

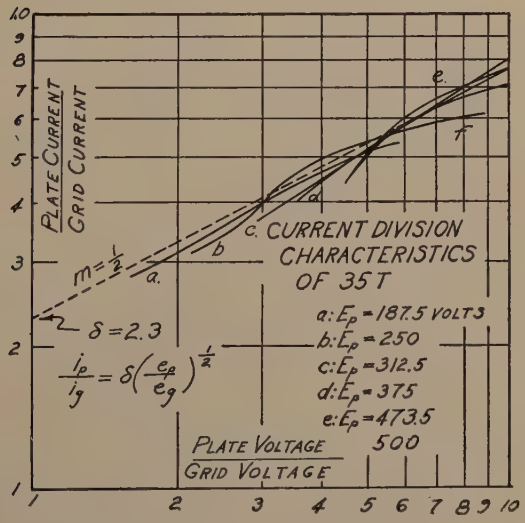


Fig. 2—i_p/i_g as a function of e_p/e_g.

curves of Fig. 1 were obtained for an Eimac 35T tube. In the operation of a class B or class C amplifier the operating conditions are almost always adjusted so that at all instants the plate voltage is greater than the grid voltages and so the only region of interest is where the

abscissa e_p/e_g is greater than unity. In this region the curves are straight when plotted on log-log paper and have a slope of 1/2. This relation was given by Tank[9] and Lange.[10] The curves nearly coincide over fairly wide ranges of e_g. Hence the functional relationship of (4) becomes

$$i_p/i_g = \delta(e_p/e_g)^{1/2} \tag{4a}$$

where δ is a constant of the tube which will be called the *current-division factor*. Fig. 2 shows a more extensive plot of the characteristics of the

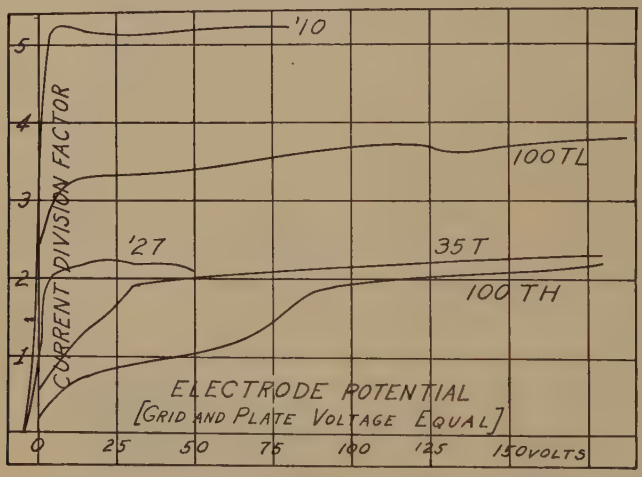


Fig. 3—δ as a function of e_p ($e_p = e_g$).

tube taken from the manufacturer's data. The form of the equation was also verified for other tubes, the slope being approximately the same for all tubes, but the current-division factor changing.

The current-division factor of a tube is most readily determined by connecting the grid and plate together and to a source of potential and measuring the ratio of the two currents. Since in this case $e_p/e_g = 1$, δ will be equal to the ratio of the two currents. Fig. 3 shows such measurements for several tubes and it is seen that δ is reasonably constant over quite a wide range, at least to the same extent as μ.

The total space current is given by the relation

[9] F. Tank, "Zur Kenntnis der Vorgänge in Elektrodenroehren," *Jahr. der Draht. Tel. und Tel.*, vol. 20, p. 80, (1922).

[10] H. Lange, "Die Stromverteilung in Dreielektrodenroehren und Ihre Bedeutung für die Messung der Voltaspannung," *Zeit. für Hochfrequenz.*, April, May, and June, (1928).

$$i_s = i_g + i_p = i_g(1 + i_p/i_g)$$

or

$$i_g = \frac{i_s}{1 + i_p/i_g}$$

$$= \frac{i_s}{1 + \delta\sqrt{e_p/e_g}}. \tag{5}$$

Now the space current is given by the well-known equation

$$i_s = K\left(e_g + \frac{e_p}{\mu}\right)^a.$$

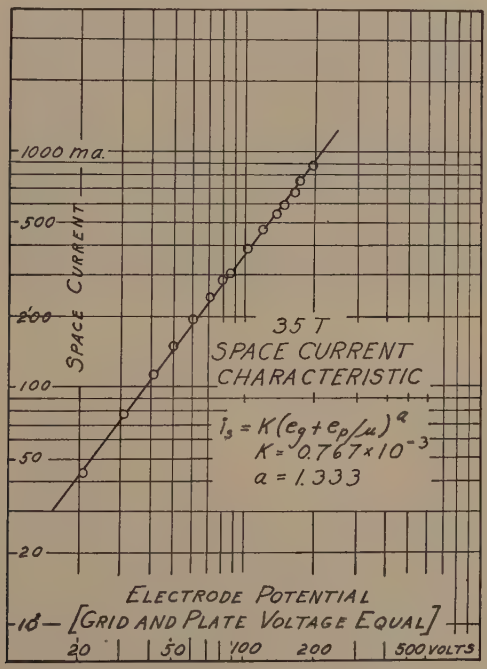


Fig. 4—i_s as a function of e_p ($e_p = e_g$).

Hence the expression for the grid current becomes

$$i_g = \frac{K\left(e + \frac{e_p}{\mu}\right)^a}{1 + \delta\sqrt{e_p/e_g}}. \tag{6}$$

Equation (6) was tested for the 35T. The constants a and K were determined in the standard way by the plot of Fig. 4 where it is found

that $a=1.333$ and $K=0.767$. From Fig. 3 a value of $\delta=2.3$ was taken and a value of $\mu=27.5$ was used. The grid currents were then computed by (6) and the results are compared with the experimental curves in Fig. 5.

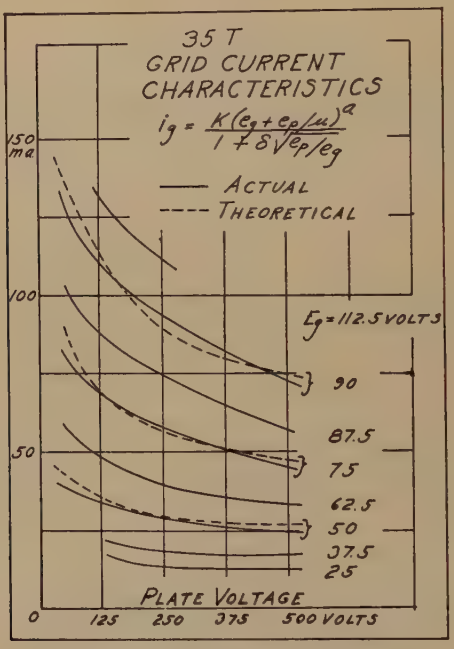


Fig. 5—i_g as a function of e_p and e_g.

The Grid-Driving-Power Equation

In order to obtain expressions for the grid-driving power and the grid loss of a class C amplifier it is necessary to know something about the form of the grid pulses. Oscillograms of some typical grid currents are shown in Fig. 6. From a study of these and other oscillograms, and also from a theoretical analysis, omitted here for the sake of brevity, it appears that the grid-current wave form tends to lie between a triangular shape and a cosine-squared wave within the angle of grid-current flow for the cases of interest where the maximum grid voltage is a fairly large proportion of the minimum plate voltage.

Both the triangular and cosine-squared wave forms are characterized by the fact that the ratio of the maximum value to the average value of current during the interval of flow is 2. This factor is called the crest factor h and is defined by the relation

$$h = \frac{i_{gm}}{i_{gav}} = \frac{i_{gm}\Theta_g}{I_{g0}\pi} \tag{7}$$

where,

i_{gm} is the maximum value of grid current

i_{gav} is the average grid current for the duration of current flow

Θ_g is one half the angle during which grid current flows and is defined by the relation

$$\Theta_g = \cos^{-1}\frac{E_c}{E_g}. \tag{8}$$

The crest factor was measured experimentally and the results are shown in Fig. 7. The region of principal interest, as will be seen later,

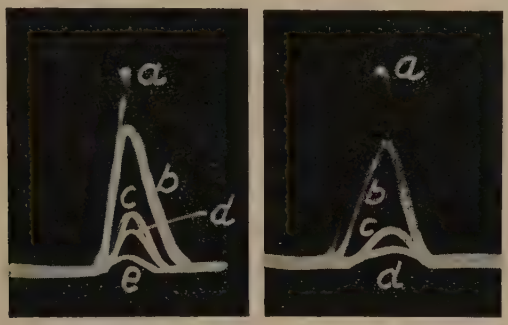


Fig. 6—Grid-current wave forms.

Class B	Class C
$E_b = 1000$ volts. $R_L = 6490$ ohms	$E_b = 1000$ volts. $R_L = 8580$ ohms
a—$E = 142$ volts	a—$E = 400$ volts
b $= 128$ volts	b $= 350$ volts
c $= 98$ volts	c $= 300$ volts
d $= 80$ volts	d $= 200$ volts
e $= 68$ volts	(E $= 232$ volts)

is for values of e_{gm}/e_{pm} between 0.6 and 1.0. Within this range h is approximately constant for given bias conditions and does not differ greatly from 2. The exact expression for the power supplied to the grid circuit from the driving source is

$$W_d = \frac{1}{\pi}\int_0^{\Theta_g} i_g E_g \cos\Theta \, d\Theta. \tag{9}$$

If a triangular wave form of grid current is assumed this integration may be performed. For this approximation

$$i_g = i_{gm}\left(1 - \frac{\Theta}{\Theta_g}\right) \qquad 0 < \Theta < \Theta_g \tag{10}$$

$$\therefore \quad W_d = \frac{E_g i_{gm}}{\pi}\int_0^{\Theta_g}\left(1 - \frac{\Theta}{\Theta_g}\right)\cos\Theta \, d\Theta$$

$$= \frac{E_g i_{gm}}{\pi}\,\frac{(1 - \cos\Theta_g)}{\Theta_g}. \tag{11}$$

For the triangular wave

$$I_{g0} = \frac{i_{gm}\Theta_g}{2\pi} .$$

Hence,

$$W_d = E_g I_{g0}\left[\frac{2(1 - \cos\Theta_g)}{\Theta_g{}^2}\right]. \tag{12}$$

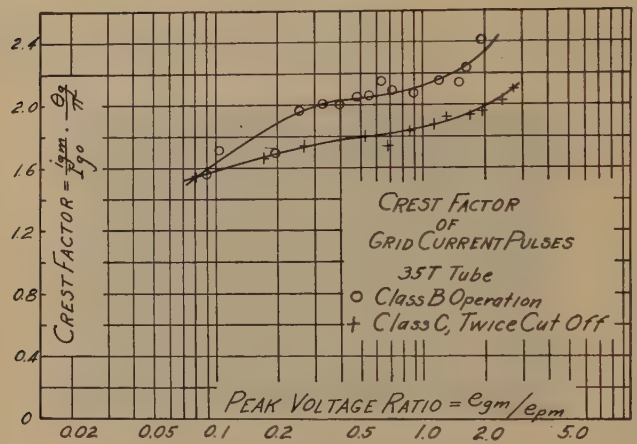


Fig. 7—h as a function of p.

This can be written

$$W_d = E_g I_{g0} t(\Theta_g) \tag{13}$$

where,

$$t(\Theta_g) = \frac{2(1 - \cos\Theta_g)}{\Theta_g{}^2} . \tag{14}$$

If a cosine-squared wave is assumed for the grid-current wave form, the grid-current approximation would be

$$i_g = \frac{i_{gm}}{2}\left[1 + \cos\frac{\pi\Theta}{\Theta_g}\right] \quad -\Theta_g < \Theta < +\Theta_g \tag{15}$$

which gives the integral

$$W_d = \frac{E_g i_{gm}}{2\pi}\int_0^{\Theta_g} \cos\Theta\left[1 + \cos\frac{\pi\Theta}{\Theta_g}\right]d\Theta$$

$$= E_g I_{g0}\left[\frac{(\pi^2)\sin\Theta_g}{\Theta_g(\pi^2 - \Theta_g{}^2)}\right]. \tag{16}$$

This can be written

$$W_d = E_g I_{g0}\, s(\Theta_g). \tag{17}$$

The two functions $t(\Theta_g)$ and $s(\Theta_g)$ are shown in Fig. 8 and it is seen that they do not differ greatly. This indicates that the grid power is

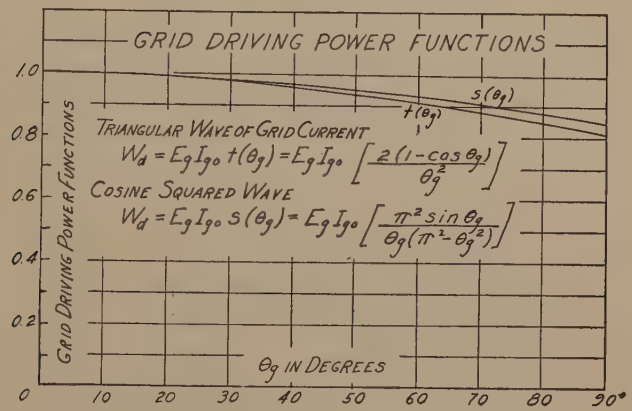


Fig. 8—$t(\Theta_g)$ and $s(\Theta_g)$ as a function of Θ_g.

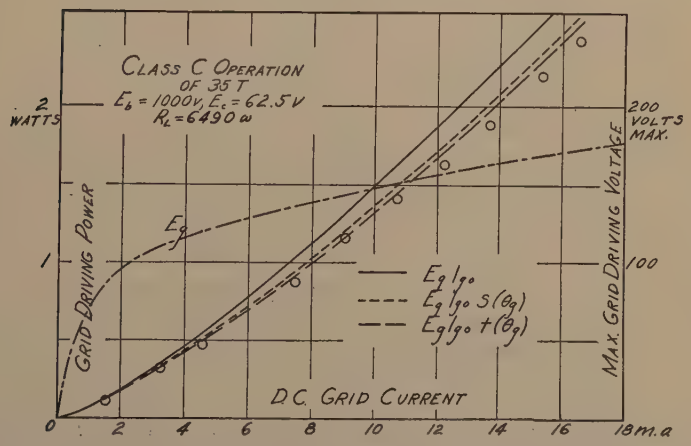


Fig. 9—Comparison of theoretical expressions for grid-driving power with experimental data. Curve E_g should be labeled e_{gm}.

not greatly affected by the wave form. Since both $t(\Theta_g)$ and $s(\Theta_g)$ are nearly equal to unity for wide ranges of Θ_g, the equation of Terman and Thomas (equation (2)) is further substantiated.

Experimental curves together with the various approximations are shown in Figs. 9 and 10, and it is seen that (13) gives the best check.

The power dissipated in the grid itself can readily be ascertained.

$$W_g = W_d - E_c I_{g0}$$
$$W_g = W_d - E_g I_{g0} \cos \Theta_g$$
$$= E_g I_{g0} [t(\Theta_g) - \cos \Theta_g] \tag{18}$$

for the triangular approximation, or

$$W_g = E_g I_{g0} [s(\Theta_g) - \cos \Theta_g] \tag{19}$$

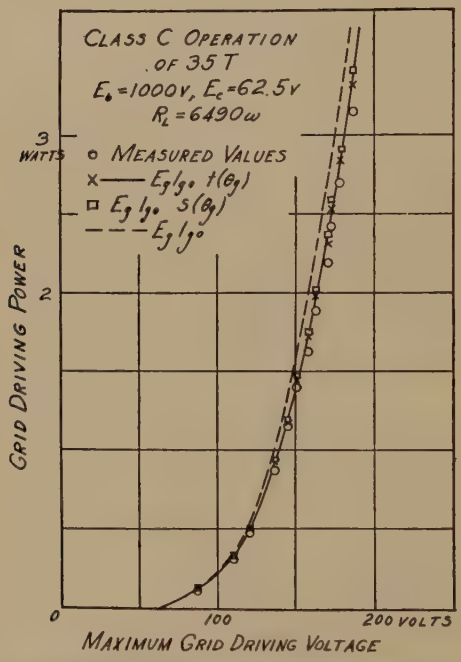


Fig. 10—Comparison of theoretical expressions for grid-driving power with experimental data.

for the cosine-squared approximation. The form of these functions is illustrated in Fig. 11. This shows that the ratio of the power lost in the grid to that lost in the entire circuit is a function of the angle of grid-current flow.

$$\frac{W_g}{W_d} = 1 - \frac{\cos \Theta_g}{t(\Theta_g)} . \tag{20}$$

For small angles of grid-current flow most of the power is dissipated in the bias source while as the angle is increased a larger proportion is dissipated on the grid.

Since the total power supplied is also given by the expression

$$W_d = \frac{E_g I_{g1}}{2}$$

it is apparent that

$$\frac{I_{g1}}{I_{g0}} = 2t(\Theta_g) \quad \text{or} \quad 2s(\Theta_g) \tag{21}$$

depending upon the wave form of the current. A check of this relation was made with a harmonic analyzer for a grid excitation at 60 cycles with the results shown in Fig. 11, which gives a good check for the theoretical function.

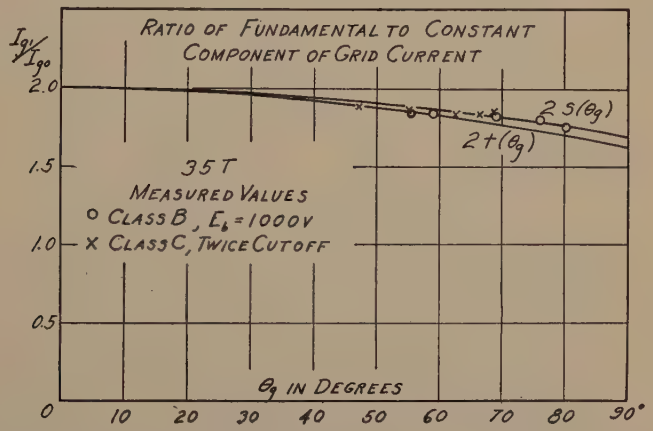


Fig. 11—Experimental check of $s(\Theta_g)$ and $t(\Theta_g)$ from current ratio I_{g1}/I_{g0}.

The Optimum Condition for Class C Operation

In the method previously proposed[1,3] for the analytical design of class C amplifiers the criterion which was established was that the tube should operate in a circuit and with applied voltages which would produce the maximum output with a given plate loss. It is here proposed to modify this criterion so that an upper limit is also set on the grid loss. If the conditions prescribed by the plate circuit alone result in a grid loss which is below the limit set, then all is well and no further steps are necessary. If, however, the grid loss is found to be excessive, then the load impedance, C bias, and grid-excitation voltage must be modified to reduce the grid loss to the allowable value, and at the same time not change the plate loss. This always means an increase in the plate resistance and a reduction in E_c and E_g. The highest output possible under these limitations should then be secured. This output will be somewhat less than that obtainable under optimum condi-

tions when the grid losses are not limited, but the reduction in output is usually only a few per cent.

It should be observed that the method of variation proposed above is not the same as that investigated by Spitzer,[4] where only the grid excitation was varied. Hence (2) does not apply.

The allowable grid loss was determined by the method suggested by Mouromtseff and Kozanowski.[8] This method makes use of the temperature at which the grid begins to emit thermally as a criterion for determining the allowable dissipation.

It has been shown that the grid-driving power and grid loss can be determined if the direct-current component of grid current, the angle of grid-current flow, and the grid-driving voltage are known. The direct-current component of grid current can be obtained by a combination of (6) and (7).

$$I_{g0} = \frac{i_{gm}\Theta_g}{h\pi} = \frac{\Theta_g}{h\pi} \frac{K\left(e_{gm} + \frac{e_{pm}}{\mu}\right)^a}{1 + \delta\sqrt{e_{pm}/e_{gm}}} \tag{22}$$

where, e_{gm} is the maximum value of grid voltage
e_{pm} is the minimum value of plate voltage.

Let $p = e_{pm}/e_{gm}$. Prince[7] has suggested a value of p equal to 1.25. This value must be assumed more accurately in the computation of grid currents than in the computation of plate currents. In the original[1] determination of optimum plate conditions p was taken equal to unity, but when the conditions computed were set up it was usually found that due to the robbing of the plate current by the grid, the actual value of p was about 1.25 and so the conditions computed were good practical values. It must be remembered that the space current diverted to the grid provides a second-order variation in the plate current, but a first-order variation in the grid current, and so approximations warranted in the plate circuit may not be accurate enough for the computation of grid currents. For example, the Child's law exponent a of (6) will be accurately determined for use in computing grid currents, but will be assumed as unity in computing plate currents.

If the ratio p is introduced, and if the value 1.8 is taken for h from Fig. 7, then (22) becomes

$$I_{g0} = \frac{\Theta_g}{1.8\pi} \frac{K\left(1 + \frac{p}{\mu}\right)^a e_{gm}{}^a}{1 + \delta\sqrt{p}} . \tag{23}$$

The value of e_{gm} must be determined from the plate operating conditions. It has been shown[1,3] that for the optimum plate condition

$$\frac{E_{LC}}{E_b} = \frac{\alpha}{\alpha + B}$$

where,

$$\alpha = \frac{(\mu + 1)R_L}{R_p}$$

$$B = \frac{\pi(1 - \cos \Theta_1)}{\Theta_1 - \sin \Theta_1 \cos \Theta_1}.$$

E_{LC} is the maximum value or amplitude of the alternating voltage across the tank circuit

E_b is the direct voltage supplied in the plate circuit

Θ_1 is one half the angle of plate-current flow.

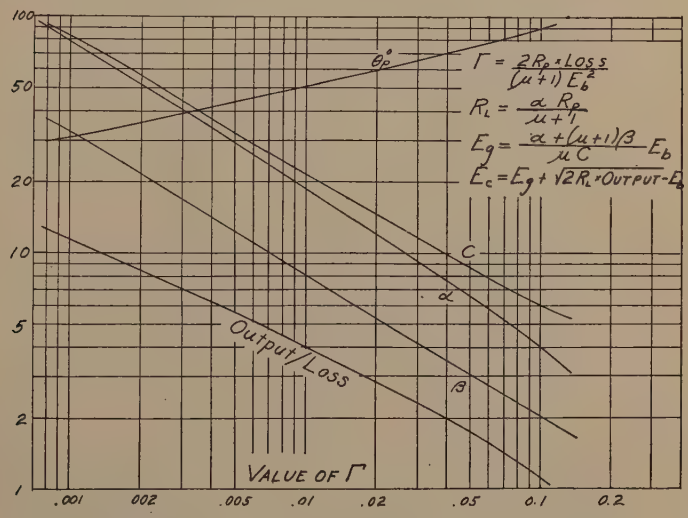


Fig. 12—Class C plate-circuit design chart.

A good approximation for the maximum grid voltage when E_{LC} is computed for the plate optimum condition is

$$e_{gm} = E_b - E_{LC}$$

$$\therefore \quad \frac{e_{gm}}{E_b} = \frac{B}{\alpha + B}. \tag{24}$$

In the previous publications[1,3] the allowable plate loss and other constants of the tube determine a numerical value designated in one case[1] by the symbol γ_3 and in the other cases[2,3] by the symbol Γ.

$$\gamma_3 = \Gamma = \frac{\text{loss} \times 2R_p}{(\mu + 1)E_b^2}.$$

From this numerical value the optimum operating conditions from the standpoint of the plate circuit may be determined by the curves which are shown in Fig. 12 and which are repeated from the previous publications.[1,3] The procedure of design will be to determine the grid loss for the optimum plate condition when the grid loss is neglected, and then apply a law for the variation of grid loss as a function of the maximum grid voltage under the condition that the load impedance and plate operating angle are simultaneously varied so as to keep the plate loss constant. When the maximum grid voltage has been reduced sufficiently to drop the grid loss to the required value the true optimum conditions will be secured.

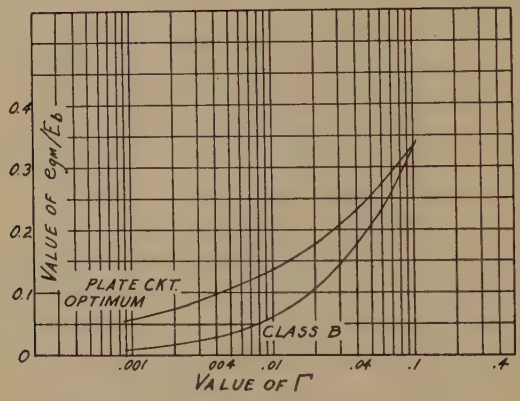


Fig. 13—e_{gm}/E_b as functions of Γ.

From Fig. 12 it is apparent that for a given value of Γ and optimum plate conditions there is a fixed value of Θ_1 and α. Hence from (24) there is a fixed value of e_{gm}/E_b. The value of e_{gm}/E_b as a function of Γ is plotted in Fig. 13.

The value of the grid excitation is given on Fig. 12 as

$$E_g = \frac{\alpha + (\mu + 1)\beta}{\mu C} E_b \tag{25}$$

where,

$$\beta = \frac{\pi}{\Theta_1 - \sin \Theta_1 \cos \Theta_1}$$

$$C = \alpha + B.$$

Now

$$E_c = E_g - e_{gm}. \tag{26}$$

Combining (8), (24), (25), and (26) the angle of grid-current flow is determined by the relation

$$\cos \Theta_g = 1 - \frac{e_{gm}}{E_g}$$

$$= 1 - \frac{\mu B}{\alpha + (\mu + 1)\beta} . \tag{27}$$

From (27) it is apparent that for any value of Γ and μ there is also a definite value of Θ_g for the optimum plate condition. The general relation

$$\Theta_g = f(\mu, \Gamma) \tag{28}$$

is shown by the family of curves of Fig. 14.

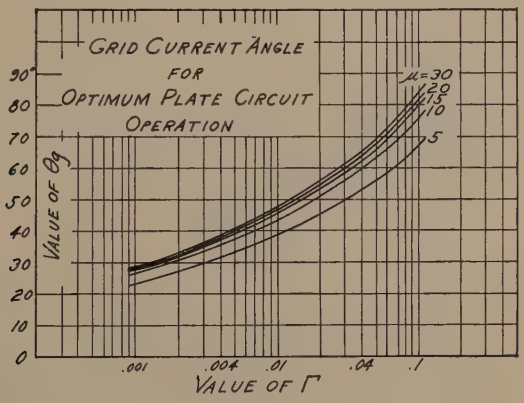


Fig. 14—Θ_g as a function of Γ and μ.

For any given tube the grid loss may now be computed for the optimum plate condition. The Γ of the tube is first determined. From Figs. 13 and 14 e_{gm}/E_b and Θ_g are determined. δ, K, and a are determined for the tube and I_{g0} is then computed by (23). $[t(\Theta_g) - \cos \Theta_g]$ is obtained from Fig. 11 and the grid loss is then computed from (19).

It is now necessary to determine how the grid loss varies with e_{gm} when the plate loss is maintained constant. If in (12) $\cos \Theta_g$ is replaced by the first three terms of its series

$$\cos \Theta_g = 1 - \frac{\Theta_g^2}{2} + \frac{\Theta_g^4}{24}$$

equation (12) becomes

$$W_d = E_g I_{g0}\left[1 - \frac{\Theta_g^2}{12}\right]. \tag{29}$$

If now the approximation is made

$$\Theta_g^2 = 2(1 - \cos \Theta_g)$$

equation (29) becomes

$$W_d = 0.833E_gI_{g0} + 0.1667E_cI_{g0}.$$

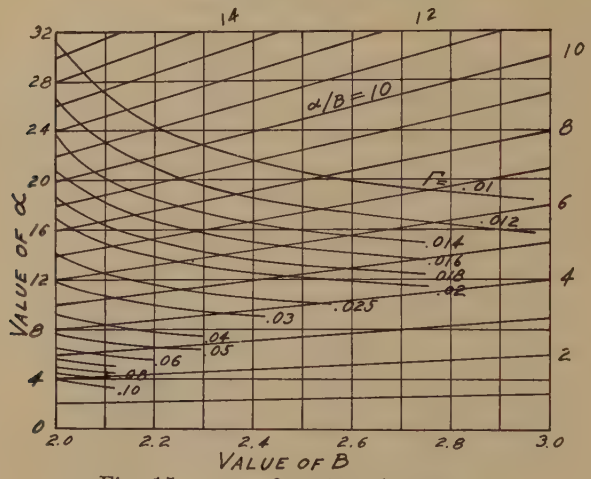


Fig. 15—α as a function of B and Γ.

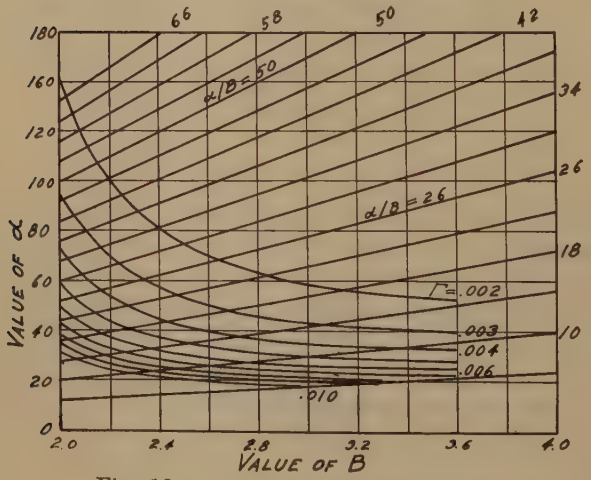


Fig. 16—α as a function of B and Γ.

Now the loss in the grid itself will be

$$W_g = W_d - E_cI_{g0} = 0.833(E_g - E_c)I_{g0}$$
$$= 0.833e_{gm}I_{g0}. \tag{30}$$

By (23) if p is kept constant, then for a given tube,

$$I_{g0} = C_1\Theta_g e_{gm}{}^a \tag{31}$$

where C_1 is a constant determined by the tube. Hence from (30)

$$W_g = 0.833 C_1 \Theta_g e_{gm}^{1+a}. \tag{32}$$

The value of Θ_g is a function of e_{gm}, Θ_g usually decreasing with increasing values of e_{gm}. It will now be shown that this relation is of the form

$$\Theta_g = C_1 e_{gm}^{b} \tag{33}$$

so that (32) becomes

$$W_g = C_2 e_{gm}^{1+a+b} = C_2 e_{gm}^{d}. \tag{34}$$

For any given loss or Γ the following relation[1,3] exists, irrespective of whether optimum conditions apply or not:

$$\Gamma = \frac{1}{A(\alpha + B)} - \frac{\alpha}{(\alpha + B)^2}. \tag{35}$$

Since A and B are functions of Θ_1, (35) can be rearranged to give such functions as

$$\alpha = f_1(\Gamma, \Theta_1) \tag{36a}$$

$$\alpha = f_2(\Gamma, B) \tag{36b}$$

$$\frac{\alpha}{B} = f_3(\Gamma, B). \tag{36c}$$

The functions of (36b) and (36c) are plotted in Figs. 15 and 16 Knowing these values one can calculate corresponding values of e_{gm} and Θ_g from (24) and (27) for each value of μ. For different values of μ and Γ curves of Θ_g versus e_{gm} were plotted on log-log paper and the curves were found to be straight lines. The slope of these lines gave the value of b in (33). This value of b as a function of Γ and μ is shown in Fig. 17.

The exponential relation of (34) for a constant plate loss was tested experimentally and the results are shown in Fig. 18. Adjustments were made to bring the plate loss and the ratio e_{pm}/e_{gm} to a particular value for each curve. The ratio e_{pm}/e_{gm} was checked by a special vacuum-tube voltmeter. Each point was actually obtained by interpolating between the values measured as the effective load resistance was varied for a fixed bias and a fixed ratio p.

From (34) the ratio by which the maximum grid voltage must be changed to reduce the grid loss to the allowable value is given by the relation

$$\frac{e_{gm1}}{e_{gm2}} = \left(\frac{W_{g1}}{W_{g2}}\right)^{1/d}. \tag{37}$$

Once the maximum grid voltage for an allowable grid loss is known the operating conditions for the tube are readily determined. Equation (24) can be rearranged to give

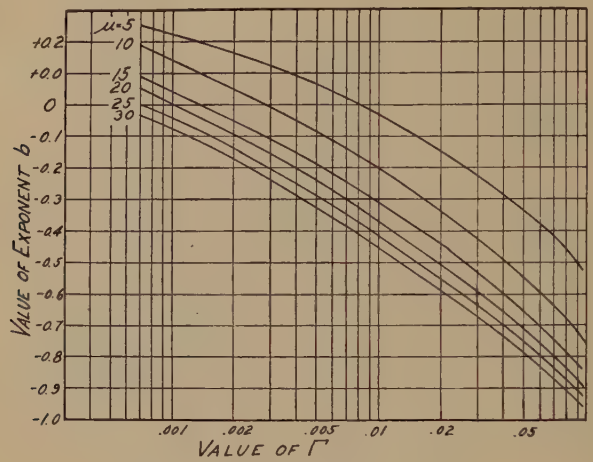


Fig. 17—b as a function of Γ and μ.

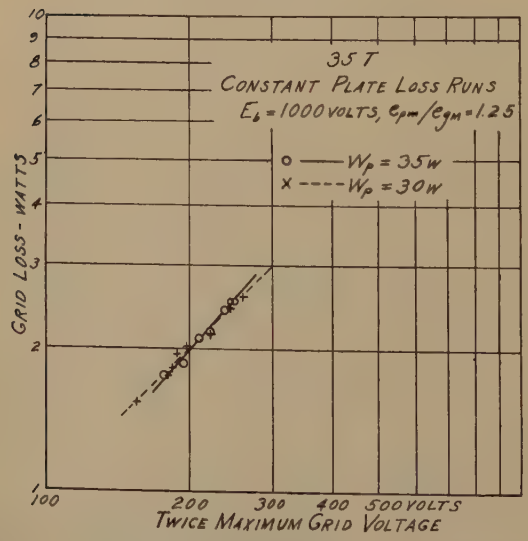


Fig. 18—Experimental check of form of equation (34).

$$\frac{\alpha}{B} = \frac{E_b}{e_{gm}} - 1. \tag{38}$$

The slope lines α/B have been drawn in Figs. 15 and 16. Hence when E_b/e_{gm} is known, α/B is determined and the value of α and B for a

given Γ may be ascertained from these figures. The output is then determined from the expression[1,3]

$$P_{\text{out}} = \frac{(\mu+1)E_b^2}{2R_p}\left[\frac{\alpha}{(\alpha+B)^2}\right]. \tag{39}$$

Examples and Experimental Checks

Although the development of the theory presented above has been somewhat long, the application of the results to an actual design problem is comparatively simple and straightforward. Illustrations of the design procedure are given below.

The 35T is a high-μ vacuum triode with a rated plate dissipation of 35 watts and nominally operated at 1200 volts. A test was made of this tube operating at 1500 volts.

The measured constants of the tube were

$$\mu = 27.5$$

$$g_m = 3700 \text{ micromhos}$$

$$R_p = 7430 \text{ ohms.}$$

Apply the plate-circuit-design procedure as follows

$$\Gamma = \frac{2R_p \times \text{loss}}{(\mu+1)E_b^2}$$

$$= \frac{2 \times 7430 \times 35}{28.5 \times 10^6}$$

$$= 0.00811.$$

Corresponding to this value of Γ one obtains from the design curves of Fig. 12, $\alpha = 21.3$, $\beta = 9.12$, $C = 24.1$, output/loss $= 4.43$. Applying these values to obtain the operating conditions there results

$$R_L = \frac{\alpha R_p}{\mu+1}$$

$$= \frac{21.3 \times 7430}{28.5}$$

$$= 5560 \text{ ohms,}$$

$$E_g = \frac{\alpha(\mu+1)\beta}{\mu c} E_b$$

$$= \frac{21.3 + 28.5 \times 9.12}{27.5 \times 24.1} 1500$$

$$= 636 \text{ volts},$$

$$\text{output} = 4.43 \times 35$$

$$= 155.3 \text{ watts},$$

$$E_c = E_g + \sqrt{2R_L \times \text{output}} - E_b$$

$$E_c = 636 + \sqrt{2 \times 7430 \times 155.3} - 1500$$

$$= 449 \text{ volts}.$$

To this point the design procedure has exactly followed that given in the previous publications.[1,3]

To calculate the grid current and grid loss, use must be made of the additional constants appearing in (6). These are the space-current coefficient K, the exponent a, and the current-division factor δ. For the 35T these are

$$K = 0.767 \times 10^{-3}$$

$$a = 1.333$$

$$\delta = 2.30.$$

The angle of grid-current flow for the operating conditions calculated above is given by

$$\Theta_g = \cos^{-1} E_c/E_g$$

$$= \cos^{-1} 449/636$$

$$= 45.0 \text{ degrees or } 0.785 \text{ radians}.$$

Using this value of Θ_g and the above constant in (22) there results

$$I_{g0} = \frac{0.785}{1.8\pi} \frac{0.767(1 + 1.25/27.5)^{1.333}}{1 + 2.3(1.25)^2} 187^{1.333}$$

in which the maximum value of grid voltage of $187 = 636 - 438$ has been used. $I_{g0} = 34.05$ milliamperes. Assuming a triangular wave of grid current the grid-loss function for 45.0 degrees is $t(\Theta_g) - \cos \Theta_g = 0.244$.

Therefore,

$$\begin{aligned} W_g &= E_g I_{g0}\, t(\Theta_g) - \cos \Theta_g \\ &= 636 \times 0.03405 \times 0.244 \\ &= 5.28 \text{ watts.} \end{aligned}$$

This value of grid loss is considerably higher than a safe operating value. Tests of the thermal emission of the grid of the 35T indicate that a maximum safe dissipation is about 2.5 watts.

Assume that a grid loss of 2.5 watts is allowable. To find the maximum grid voltage corresponding to this loss the voltage must be scaled off according to the relation of (37). From the curves of Fig. 17 it is seen that the value of the exponent b corresponding to a Γ of 0.00811 and a μ of 27.5 is -0.385. Therefore,

$$\begin{aligned} d &= 1 + a + b \\ &= 1 + 1.333 - 0.385 \\ &= 1.948 \end{aligned}$$

and

$$1/d = 0.513.$$

Then the maximum grid voltage corresponding to a grid loss of 2.5 watts is given by

$$\begin{aligned} e_{gm} &= 187(2.5/5.28)^{0.513} \\ &= 127.6 \text{ volts.} \end{aligned}$$

Hence,

$$\begin{aligned} \alpha/B &= E_b/e_{gm} - 1 \\ &= 1500/127.6 - 1 \\ &= 10.75. \end{aligned}$$

Interpolating between the curves of Fig. 16 to find the value of α and B corresponding to a slope of 10.75 and a Γ of 0.00811 it is found that $\alpha = 25.6$ and $B = 2.39$. The corresponding value of β is 2.09.

Corresponding to the values of α and B determined above the following operating conditions are obtained:

$$\begin{aligned} R_L &= \frac{\alpha R_p}{\mu + 1} \\ &= \frac{25.6 \times 7430}{28.5} \\ &= 6680 \text{ ohms.} \end{aligned}$$

By (39)

$$\text{output} = \frac{28.5 \times 2.25 \times 10^6}{14860} \frac{25.6}{282}$$

$$= 137 \text{ watts.}$$

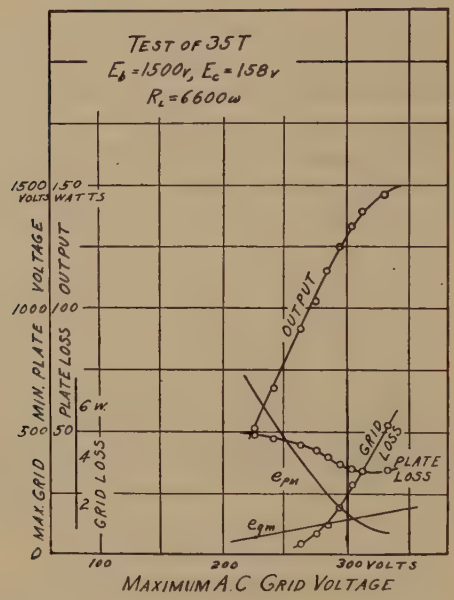


Fig. 19—Experimental verification of grid design procedure on Eimac 35T.

By (25)

$$E_g = \frac{25.6 + 28.5 \times 4.1}{27.5 \times 28} 1500$$

$$= 280 \text{ volts}$$

and

$$E_c = E_g - e_{gm}$$

$$= 280 - 128$$

$$= 152 \text{ volts.}$$

The above operating conditions were set up as accurately as adjustments upon the apparatus would permit and the actual and theoretical performances were compared. The results of this test are shown in Fig. 19. It will be observed that the ratio of minimum plate voltage to the maximum grid voltage is about 1.25 at a grid-driving voltage of 293 volts. At this voltage the plate loss is 35.3 watts and the grid loss 2.1

watts. The output is 128 watts. Comparing the measured and predicted values it is seen that they are in fair agreement. The computed maximum grid voltage always comes out distinctly less than the measured value at which the maximum grid and minimum plate voltage are actually equal. The output is less than expected because of the large fraction of the space current diverted to the grid of the tube.

The 100TH is a high-voltage vacuum triode with a plate dissipation of 100 watts and built to operate with 3000 volts on the plate. Limitations of the laboratory high-voltage supply did not permit the operation of the tube at voltages higher than 2000 volts. A calculation will be made upon the 100TH for operation at 2000 volts.

$$\mu = 30$$

$$g_m = 3800 \text{ micromhos}$$

$$R_p = 7980 \text{ ohms.}$$

Apply the optimum plate-circuit design procedure

$$\Gamma = \frac{2R_p \times \text{loss}}{(\mu + 1)E_b^2}$$

$$= \frac{15960 \times 100}{31 \times 4 \times 10^6}$$

$$= 0.01288.$$

Corresponding to this value of Γ one obtains from the design curves $\alpha = 6.6$, $\beta = 6.87$, $C = 18.5$, output/loss $= 3.5$. Applying these values to obtain the operating conditions there results

$$R_L = \frac{\alpha R_p}{(\mu + 1)}$$

$$= \frac{16.6 \times 7980}{(\mu + 1)}$$

$$= 4270 \text{ ohms}$$

$$\text{output} = 3.5 \times 100 = 350 \text{ watts}$$

$$E_g = \frac{\alpha + (\mu + 1)\beta}{\mu C} E_b$$

$$= \frac{16.6 + 31 \times 6.87}{30 \times 18.5} 2000$$

$$= 829 \text{ volts}$$

$$\begin{aligned} E_c &= E_g + \sqrt{2R_L \times \text{output}} - E_b \\ &= 829 + \sqrt{8450 \times 350} - 2000 \\ &= 559 \text{ volts} \\ e_{gm} &= E_g - E_c \\ &= 829 - 559 \\ &= 270 \text{ volts.} \end{aligned}$$

To this point the design procedure has exactly followed that given in the previous publications.[1,3]

The above operating conditions are seen to be rather extreme because of the very high grid-driving voltage and grid bias called for. The power amplification under these conditions is bound to be low.

For the 100TH

$$\begin{aligned} K &= 0.580 \times 10^{-3} \\ a &= 1.405 \\ \delta &= 2.30. \end{aligned}$$

The angle of grid-current flow for the operating conditions calculated above is given by

$$\begin{aligned} \Theta_g &= \cos^{-1} E_c/E_g \\ &= \cos^{-1} 559/829 \\ &= 47.6 \text{ degrees or } 0.83 \text{ radians.} \end{aligned}$$

Using this value of Θ_g and the above constants in (22) there results

$$\begin{aligned} I_{g0} &= \frac{0.83}{1.8\pi} \frac{0.58(1 + 1.25/30)^{1.405}}{1 + 2.3(1.25)^{1/2}} 270^{1.405} \\ &= 65.6 \text{ milliamperes.} \end{aligned}$$

Assuming a triangular wave of grid current the grid-loss function corresponding to 47.6 degrees is $t(\Theta_g) - \cos \Theta_g = 0.270$. Therefore,

$$\begin{aligned} W_g &= E_g I_{g0}[t(\Theta_g) - \cos \Theta_g] \\ &= 829 \times 0.0656 \times 0.270 \\ &= 14.69 \text{ watts.} \end{aligned}$$

This value of grid loss is much higher than is safe to operate. Tests of the thermal emission of the 100TH indicate that the grid dissipation should not be greatly in excess of 5 watts.

Assume that a grid loss of 5 watts is allowable. To find the maximum

grid voltage corresponding to this loss the voltage must be scaled off according to the relation of (37). From the curves of Fig. 17 it is seen that the value of the exponent b corresponding to a Γ of 0.01288 and a μ of 30 is -0.510. Therefore

$$\begin{aligned} d &= 1 + a + b \\ &= 1 + 1.405 - 0.510 \\ &= 1.895 \end{aligned}$$

and

$$1/d = 0.528.$$

Then the maximum grid voltage corresponding to the loss of 5 watts is given by

$$\begin{aligned} e_{gm} &= 270(5/14.69)^{0.528} \\ &= 152.2 \text{ volts.} \end{aligned}$$

Hence,

$$\begin{aligned} \alpha/B &= E_b/e_{gm} - 1 \\ &= 2000/152.2 - 1 \\ &= 12.15. \end{aligned}$$

Interpolating between the curves of Fig. 16 to find the value of α and B corresponding to a slope of 12.15 and a loss function of 0.01288 it is found that $\alpha = 24.5$ and $B = 2.018$. The corresponding value of β is 2.09. It will be noted that this gives an operating condition which is almost the class B case.

Corresponding to the values of α and B determined above the following operating conditions were obtained:

$$R_L = \frac{\alpha R_p}{\mu + 1}$$

$$\begin{aligned} R_L &= \frac{24.5 \times 7980}{31} \\ &= 6310 \text{ ohms.} \end{aligned}$$

By (39)

$$\begin{aligned} \text{output} &= \frac{31 \times 4 \times 10^6}{2 \times 7980} \frac{24.5}{26.5^2} \\ &= 271 \text{ watts.} \end{aligned}$$

By (25)

$$E_g = \frac{24.5 + 31 \times 2.09}{30 \times 26.5} 2000$$

$$= 224 \text{ volts}$$

$$E_c = E_g - e_{gm}$$

$$= 224 - 152$$

$$= 72 \text{ volts.}$$

The above operating conditions were set up as closely as adjustments upon the apparatus would permit and the actual and theoretical

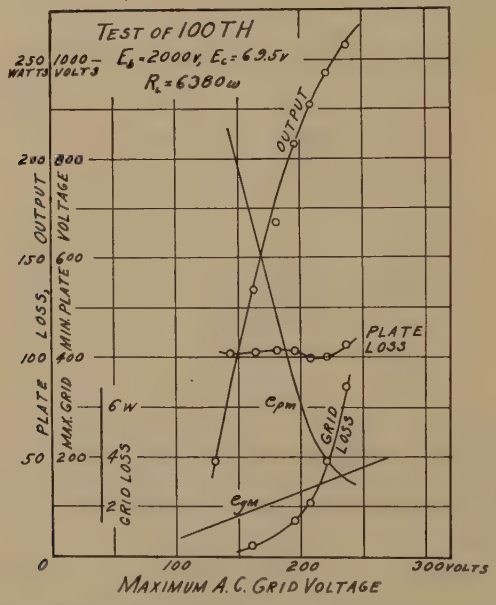


Fig. 20—Experimental verification of grid design procedure on Eimac 100TH.

performances were compared. The results of this test on the 100TH are shown in the curves of Fig. 20. It will be observed that the ratio of the minimum plate voltage to the maximum grid voltage at a voltage of 224 volts is about 1.25. At this voltage the plate loss is 101 watts and the grid loss is 4.3 watts. The output is 246 watts. Comparing the predicted with the measured results we have the following:

Predicted	Measured
$E_g = 224$ volts	$E_g = 224$ volts
plate loss = 100 watts	plate loss = 101 watts
$W_g = 5.0$ watts	$W_g = 4.3$ watts
$P_{out} = 271$ watts	$P_{out} = 246$ watts.

The predicted and measured values are in fair agreement. The difference between the measured and predicted values of output can be ascribed principally to the large percentage of space current which the grid robs from the plate at this voltage. At the predicted voltage the direct-current component of grid current is approximately one fourth of the direct-current component of plate current. It will be noticed that at the predicted condition a change of only 8 volts in driving voltage, about a 4 per cent change, produces a change of 1 watt in grid loss or over a 20 per cent change. It will be noticed further that the predicted voltage occurs at the point at which the plate loss begins to increase which is in accord with the theory.

In both of the cases tested above the modification of the optimum plate-circuit conditions produces a true over-all optimum operating condition. A maximum output is achieved without exceeding either the allowable plate or grid loss. Design on this basis also achieves the result of producing a high power amplification. In both cases the power amplification is over 20 to 1.

The method outlined above is seen to be fairly accurate. Since slight adjustments of bias, load, grid-driving voltage, and so on can be made at the predicted condition which will compensate for any departures from the desired operating conditions, it is seen that the method is a practical one. The design method is quite straightforward and does not require any guesswork nor the application of experience factors. Although the method properly has been devised for cases in which there is slight secondary emission the method will give operating conditions with a factor of safety if applied to tubes with secondary emission as though there were none because the effect of secondary emission is always to reduce the grid-loss and grid-driving power.

The examples have been applied to class C amplifiers with constant grid excitation and plate voltage. The methods could also be applied to either the case where the plate voltage is varied (modulated amplifier) or the case where the grid excitation is varied (linear amplifier).

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

CHARACTERISTICS OF THE IONOSPHERE AT WASHINGTON, D. C., MARCH, 1938*

By

T. R. Gilliland, S. S. Kirby, and N. Smith

(National Bureau of Standards, Washington, D. C.)

DATA on the critical frequencies and virtual heights of the ionosphere layers are given for March, 1938, in Fig. 1. Fig. 2 gives the maximum frequencies which could be used for radio communication in latitudes approximately that of Washington, calculated from the data of Fig. 1.

Figs. 1 and 2 show that the night critical frequencies and maximum usable frequencies for March were considerably greater than those for February. The midday F_2-layer critical frequencies, however, were somewhat lower but remained high more hours per day than during February. This flattening out of the diurnal characteristics is expected to continue as summer is approached. It should be noted that although an appreciable separation between h_{F_1} and h_{F_2} is shown in Fig. 1, the transition between the two layers was indicated only by a slight change of slope of the frequency—virtual-height curve. Since h_{F_1} and h_{F_2} as plotted in Fig. 1 are minimum virtual heights they are observed at different frequencies, and the separation shown results largely from the continuous increase of heights with frequency.

The following critical frequencies for March, 1938, were less than those for the corresponding hours in March, 1937, by approximately the following amounts: noon f_{F_2}, 60 kilocycles; midnight f_F, 100 kilocycles; diurnal minimum (0530 local time), 300 kilocycles; noon f_E, 80 kilocycles.

Severe ionosphere storms occurred on March 22 and 23. More moderate disturbances were indicated on other days. An interesting point concerning the structure of the F region was illustrated by the ionosphere storms of March 22 and 23, as well as by storms previously reported[1] for winter ionosphere conditions. On ionospherically quiet days the F_1 and F_2 layers separate appreciably only during the summer day

* Decimal classification: R113.61. Original manuscript received by the Institute, April 9, 1938. This is one of a series of reports on the characteristics of the ionosphere at Washington, D. C. For earlier publications on this subject see Proc. I.R.E., vol. 25, pp. 823–840; July, (1937), and a series of monthly reports beginning in Proc. I.R.E., vol. 25, pp. 1174–1191; September, (1937). Publication approved by the Director of the National Bureau of Standards of the U. S. Department of Commerce.

[1] *Phys. Rev.*, vol. 48, p. 849, (1935).

when the F_2 layer appears to rise away from the F_1 layer. During the normal winter day the F_2-layer height remains low so that the F_2 layer merges with the F_1 layer. During ionosphere storms, however, the F_2 layer rises so that stratification of the F layer is observed even during the winter day. At night, however, a different situation is found. No stratification into F_1 and F_2 layers is ever observed at night even during an ionosphere storm when the F-layer heights are abnormally great.

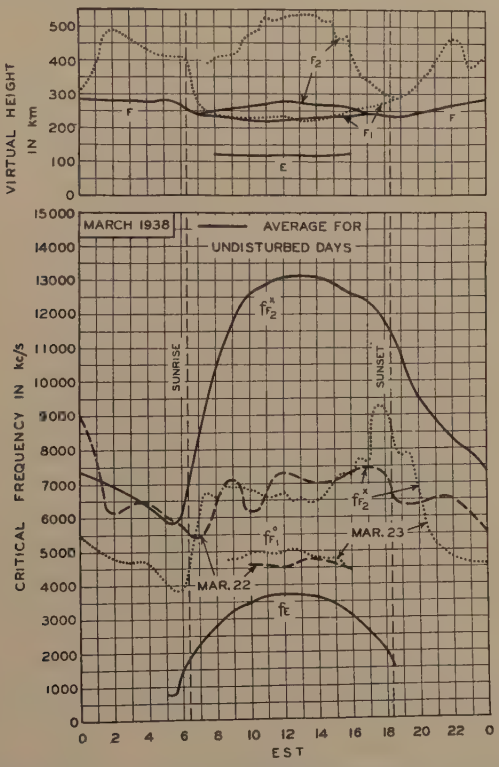


Fig. 1—Virtual heights and critical frequencies of the E, F, F_1, and F_2 layers of the ionosphere for March, 1938. The solid-line graphs represent the average for undisturbed days. The dashed and dotted graphs represent values for the ionosphere storm days of March 22 and 23, respectively.

These phenomena indicate that during the daytime at all seasons both F_1 and F_2 layers exist as two types of ionization or ionization from two different sources, requiring only physical separation to make them manifest. At night, on the other hand, it appears that the F region consists of only one layer involving only one type of ionization.

The ionosphere storms during March were more severe than those in February but not as severe as those during January. With the ap-

proach of spring transition conditions the ionosphere appeared to be more easily disturbed than during the winter. The ionosphere storms during March, 1938, are shown in Table I approximately in the order of their severity.

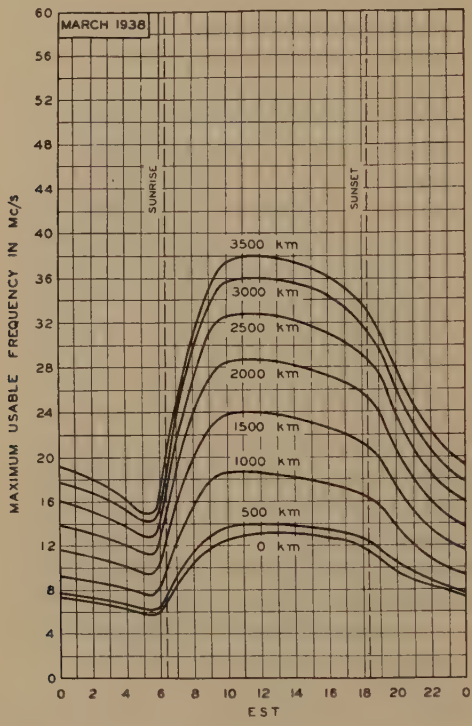


Fig. 2—Maximum usable frequencies for latitude of Washington, average for undisturbed days of March, 1938. Time to be used is local time where the waves are reflected from the ionosphere layer.

TABLE I

Date 00–24 E.S.T.	h_F before sunrise km	Minimum f_F^x during day (before sunrise) kc	Noon $f_{F_2}^x$ kc	Magnetic character[1]	
				00–12 G.M.T.	12–24 G.M.T.
March 23	436	3950	6,700	1.0	0.9
March 22	364	5450	7,000	1.4	0.9
March 24	362	4700	approximately 10,000	1.4	0.2
March 25	318	4600	8,900	0.4	0.4
March 27	318	5100	9,900	0.1	0.0
March 26	296	5400	10,500	0.9	0.4
March 5	318	5100	near average	0.8	1.1
Average of undisturbed days	281	5920	13,040	0.2	0.2

[1] American character figure, compiled by the Department of Terrestrial Magnetism, Carnegie Institution of Washington, from data supplied by their two observatories and five observatories of the United States Coast and Geodetic Survey.

Table II shows the number of hours $f_F{}^2$ differed from the average for the undisturbed days of March by more than the given percentages. It should be noted that nearly all values more than 20 per cent below the monthly average occurred on the disturbed days listed in Table I.

TABLE II

Percent	−50	−40	−30	−20	−10	−0	+0	+10	+20
For 345 hours of observations between 2000 and 0700 local time									
Number of hours	0	3	11	31	80	202	143	30	3
Disturbed hours	0	3	11	27	45	73	6	0	0
Undisturbed hours	0	0	0	4	35	129	137	30	3
For 60 hours of observations on Wednesdays between 0800 and 1900 local time									
Number of hours	2	8	10	12	12	36	24	0	0
Disturbed hours	2	8	10	12	12	12	0	0	0
Undisturbed hours	0	0	0	0	0	24	24	0	0

Sudden disturbances of the ionosphere at Washington during March were marked by the radio fade-outs listed in Table III.

TABLE III

Date	Beginning of fade-out[1]	Beginning of recovery	Recovery complete	Location of transmitter	Remarks	Minimum relative intensity[2]
March 9	1803	1818	1830	Ohio		0.1
March 15	1408	1423	1430	Ohio, Mass., D.C.		0.0
March 15	1600	1615	1640	Ohio, Mass., D.C.		0.0
March 16	2144	2155	2220	Ohio, Mass., D.C.	Ter. mag. pulse[3]	0.01
March 23	1755	1802	1820	Ohio, D.C.		0.01
March 24	1608	1628	1640	Ohio, Mass., D.C.		0.0

[1] All times G.M.T.
[2] In terms of received wave intensity from W8XAL, 6060 kilocycles, distance 650 kilometers.
[3] Terrestrial magnetic pulse observed on the Cheltenham magnetograms of the United States Coast and Geodetic Survey.

Out of 719 hours of observations during March, strong sporadic *E* reflections occurred above 4500 kilocycles during only two hours.

Note: The National Bureau of Standards broadcasts current ionosphere data and maximum usable frequencies, each Wednesday, by radiotelephone from its station WWV, in accordance with the following schedule: 1:30 P.M., E.S.T., 10 megacycles; 1:40 P.M., E.S.T., 5 megacycles; 1:50 P.M., E.S.T., 20 megacycles.

Proceedings of the Institute of Radio Engineers
Volume 26, Number 5 *May, 1938*

BOOK REVIEW

Television Engineering, by J. C. Wilson. Published by Sir Isaac Pitman and Sons, Ltd. Available through Pitman Publishing Company, 2 W. 45th St., New York, N. Y. 492 pages. Price, $10.00.

It is seldom that this reviewer can so wholeheartedly endorse a technical book on television as he can this one. Not only will engineers and scientists interested in television want a copy of this book, but patent attorneys, coming in contact with this new field, will find it a valuable volume. Mr. Wilson was a member of the Patents Department of the Baird Television Company in England at the time he wrote "Television Engineering."

Following a very brief historical chapter the author very fittingly discusses optics and the eye. We wish he had dealt at even greater length with this interesting subject.

Scanning methods and devices are next treated and the multitudinous methods of scanning, after being divided into focal plane and nonfocal plane systems, are conveniently listed, with their inventors, and even with the patent numbers, on large-sized charts. How often have television engineers been approached by would-be inventors who desire an opinion on the novelty, for instance, of some multiple-spiral television-scanner idea! In the future it will be necessary only to turn to one of these charts and look for the prototype.

Chapters are devoted to the analysis of finite-aperture-scanning methods, the frequency characteristics of various types of apertures being derived. A discussion of the physical basis of photoelectricity leads into a description of various types of photocells, their color responses, and secondary emission multipliers.

Next follows a chapter on amplifiers. A television coworker of mine, who is an amplifier expert, says that this is the best section of the book. Transmission engineers will be interested in a discussion of video-frequency transmission lines, filters, and corrective networks of the constant-resistance type. The author treats in an original method resonant circuits.

The modulation of light by Kerr cells and supersonic cells is described and their optical efficiencies examined. The properties of fluorescent materials are given. This is followed by a very clear dissertation on electron optics, in which both electric and magnetic deflection and focusing is discussed. There is also a discussion of types of deflecting circuits, which is rather too brief.

Thirty-three pages are devoted to special television methods; these include color, stereoscopic, and intermediate film transmission. Electronic camera tubes of the cumulative and noncumulative type are shown and described.

In a chapter entitled "Modern Television Equipment" there are photographs and descriptions of many well-known systems of the type operating here and abroad.

Mr. Wilson concludes his book with a discussion on the effect of frequency-band width on image detail. We feel that his use of the word "bildpunkte" for "picture elements" is rather unfortunate. The book contains commendably few errors.

The pages at the end of each chapter, giving a complete bibliography (the references being both to technical literature and to patents), form a valuable feature, especially to those working in the patent field.

*A. F. Murray

* Philco Radio and Television Corporation, Philadelphia, Pennsylvania.

CONTRIBUTORS TO THIS ISSUE

Bachman, Charles H.: Born 1908 at Cedar Falls, Iowa. Received B.S. degree in electrical engineering, 1932; M.S. degree in physics, 1933; Ph.D. degree in physics, 1935, Iowa State College. Hygrade Sylvania Corporation, 1935 to date. Member Sigma Xi, Tau Beta Pi, and American Physical Society. Student member, Institute of Radio Engineers, 1932; Associate member, 1935.

Blanchard, Julian: Born November 10, 1885, at Hertford, North Carolina. Received A.B. degree, Trinity College (now Duke University), 1905; A.M. degree, Columbia University, 1909; Ph.D. degree in physics, 1917. Instructor in applied mathematics, Trinity College, 1905–1906; assistant in physics, Columbia University, 1906–1909; professor of engineering, Trinity College, 1909–1912; research assistant, department of physics, Columbia University, 1912–1915. Physicist, research laboratory, Eastman Kodak Company, 1915–1917; research and development work on vacuum tubes, engineering department, Western Electric Company, 1917–1925; research department, Bell Telephone Laboratories, Inc., 1925 to date. Member, Sigma Xi, Phi Beta Kappa, and Optical Society of America. Nonmember, Institute of Radio Engineers.

Burrows, Charles R.: Born June 21, 1902, at Detroit, Michigan. Received B.S.E. degree in electrical engineering, University of Michigan, 1924; M.A. degree in physics, Columbia University, 1927; E.E. degree, University of Michigan, 1935. Research assistant, University of Michigan, 1922–1923. Research department, Western Electric Company, 1924–1925; research department, Bell Telephone Laboratories, Inc., 1925 to date. Associate member, Sigma Xi. Member, American Institute of Electrical Engineers. Associate member, Institute of Radio Engineers, 1924; Member, 1938.

Carnahan, Chalon W.: Born July 4, 1907, at Berkeley, California. Received B.A. degree in physics, Stanford University, 1927; M.A. degree, 1931. Instructor, Fresno State College, 1927–1930. Television Laboratories, Ltd., 1931–1933; research department, Hygrade Sylvania Corporation, 1933 to date. Associate member, Institute of Radio Engineers, 1934.

Decino, Alfred: Born June 23, 1907, at Pueblo, Colorado. Received B.S. degree in electrical engineering, University of Colorado, 1928. Member of technical staff, Bell Telephone Laboratories, Inc., 1928 to date. Associate member, Institute of Radio Engineers, 1929.

Everitt, William Littell: Born April 14, 1900, at Baltimore, Maryland. Received E.E. degree, Cornell University, 1922; M.S. degree, University of Michigan, 1926; Ph.D. degree, Ohio State University, 1933. Instructor in electrical engineering, Cornell University, 1920–1922; engineer in charge of laboratory and P.B.X. development, North Electrical Manufacturing Company, 1922–1924; instructor in electrical engineering, University of Michigan, 1924–1926; assistant professor of electrical engineering in charge of communication courses, Ohio State University, 1926–1929; associate professor, 1929–1934; professor, 1934 to date. Department of development and research, American Telephone and Telegraph Company, summers, 1925–1930, inclusive. Associate member, Institute of Radio Engineers, 1925; Member, 1929; Fellow, 1938.

Gilliland, T. R.: See PROCEEDINGS for January, 1938.

Hunt, Loyd E.: Born November 7, 1901, at Boise, Idaho. Received A.B. degree in physics, Reed College, 1927. Commercial wireless operator and ship wireless inspector, 1918–1926; radio instructor, Oregon Institute of Technology, 1926–1927; director, radioelectric division, United Y.M.C.A. schools, Seattle, 1927–1929; member of technical staff, Bell Telephone Laboratories, Inc., 1929 to date. Member, Institute of Radio Engineers, 1933.

Kirby, S. S.: See PROCEEDINGS for January, 1938.

Leeds, Laurance M.: Born October 10, 1906, at Tulsa, Oklahoma. Received B.S. degree in electrical engineering, Rutgers University, 1934. Amateur, 1917; Manhattan Electrical Supply Company, 1922; Peck Radio Corporation, 1922; marine inspector, Radio Corporation of America, 1922; Westinghouse Electric and Manufacturing Company, station WBZ, 1923; marine operator, 1923–1926; radio inspector, Tropical Radio Telegraph Company, 1926–1927; New Jersey Bell Telephone Company, 1927–1929; transatlantic service, American Telephone and Telegraph Company, 1929–1930; radio transmitter engineering department, General Electric Company, 1934 to date. Associate member, Institute of Radio Engineers, 1930.

Loughren, Arthur V.: Born September 15, 1902, at Rensselaer, New York. Received A.B. degree, 1923; E.E. degree, 1925, Columbia University. Vacuum tube engineering department, General Electric Company, 1925–1927; radio engineering department, 1927–1929; engineering department, RCA Victor and RCA Manufacturing Company, Inc., 1930–1934; radio receiver engineering section, General Electric Company, 1934–1936; Hazeltine Service Corporation 1936 to date. Member, Sigma Xi and Tau Beta Pi. Associate member, Institute of Radio Engineers, 1924; Member, 1929.

Roder, Hans: Born September 27, 1899, at Mengersreuth, Bavaria, Germany. Received M.S. degree in electrical engineering, Institute of Technology, Munich, 1923. Development and research of transmitting apparatus, Telefunken Gesellschaft für drahtlose Telegraphie, 1923–1929; radio transmitter department, General Electric Company, 1930–1933; in charge of advanced development, radio receiver engineering section 1933–1937; general engineering laboratory, 1937 to date. Member, Institute of Radio Engineers, 1929.

Smith, Newbern: See PROCEEDINGS for January, 1938.

Spangenberg, Karl. Born April 9, 1910, at Cleveland, Ohio. Received B.S. degree in electrical engineering, 1932; M.S. degree, 1933 Case School of Applied Science; Ph.D. degree, 1937 Ohio State University. Radio engineer, station WHK, 1934; Instructor in electrical engineering, Rose Polytechnic Institute, 1935–1936. Instructor in electrical engineering, Stanford University, 1937 to date. Member Sigma Xi and American Institute of Electrical Engineers. Associate member, Institute of Radio Engineers, 1934.

Wheeler, Harold Alden: Born May 10, 1903, at St. Paul, Minnesota. Received B.S. degree in physics, George Washington University, 1925. Postgraduate studies, physics department, Johns Hopkins University, 1925–1928; lecturer, 1926–1927. Assistant, Radio Section, Bureau of Standards, 1921–1922; assistant to Professor Hazeltine, 1923; engineer, Hazeltine Corporation and Hazeltine Service Corporation, 1924 to date. Member, Sigma Xi. Associate member, Institute of Radio Engineers, 1927; Member, 1928; Fellow, 1935.